

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 09:22:57 ; Search time 14 Seconds
(without alignments)
487.580 Million cell updates/sec

Title: us-09-897-322-2
Perfect score: 1268
Sequence: 1 MERCPSLGVTLYALVVVLG.....QRGWIPIQYIISCKKSC 232

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2.6/prodata/1/iaa/5A COMB.pep:*
2: /cgn2.6/prodata/1/iaa/5B COMB.pep:*
3: /cgn2.6/prodata/1/iaa/6A COMB.pep:*
4: /cgn2.6/prodata/1/iaa/6B COMB.pep:*
5: /cgn2.6/prodata/1/iaa/PCTUS COMB.pep:*
6: /cgn2.6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1268	100.0	232	2	US-08-485-721-2
2	1268	100.0	232	2	US-08-392-935-2
3	1268	100.0	232	3	US-08-897-236-2
4	1268	100.0	232	4	US-09-167-874-2
5	1268	100.0	232	5	PCT-US93-08326-2
6	1258	99.2	232	3	US-08-897-236-11
7	1258	99.2	232	4	US-09-167-874-11
8	1195	94.2	449	3	US-08-897-236-23
9	949.5	74.9	222	2	US-08-485-721-9
10	949.5	74.9	222	2	US-08-392-935-9
11	949.5	74.9	222	5	PCT-US93-08326-9
12	503	39.7	87	2	US-08-485-721-11
13	503	39.7	87	2	US-08-392-935-11
14	503	39.7	87	5	PCT-US93-08326-11
15	93.5	7.4	372	3	US-08-586-165-5
16	92	7.3	538	4	US-09-309-572-12
17	92	7.3	1737	4	US-09-309-572-13
18	90.5	7.1	483	1	US-08-194-338-7
19	90.5	7.1	623	4	US-09-347-801-4
20	88.5	7.0	371	3	US-08-586-165-3
21	88.5	7.0	979	3	US-08-870-529-2
22	87	6.9	14	1	US-08-297-633A-3
23	87	6.9	14	2	US-08-485-721-3
24	87	6.9	14	2	US-08-392-935-3
25	87	6.9	14	3	US-08-897-236-3
26	87	6.9	14	4	US-09-167-874-3
27	87	6.9	14	5	PCT-US93-08325-3

28	87	6.9	14	5	PCT-US93-08326-3	Sequence 3, Appli
29	86	6.8	625	4	US-09-347-801-18	Sequence 18, Appl
30	85.5	6.7	315	4	US-08-957-351-27	Sequence 27, Appl
31	85	6.7	1621	1	US-08-242-677-2	Sequence 2, Appli
32	84	6.6	1593	4	US-08-628-829-4	Sequence 4, Appli
33	83	6.5	538	4	US-09-370-368-9	Sequence 9, Appli
34	79	6.2	11877	4	US-09-105-537-6	Sequence 6, Appli
35	76	6.0	14	1	US-08-297-633A-5	Sequence 5, Appli
36	76	6.0	14	2	US-08-485-721-5	Sequence 5, Appli
37	76	6.0	14	2	US-08-392-935-5	Sequence 5, Appli
38	76	6.0	14	3	US-08-897-236-5	Sequence 5, Appli
39	76	6.0	14	4	US-09-167-874-5	Sequence 5, Appli
40	76	6.0	14	5	PCT-US93-08325-5	Sequence 5, Appli
41	76	6.0	14	5	PCT-US93-08326-5	Sequence 5, Appli
42	75.5	6.0	6095	4	US-09-144-085-2	Sequence 2, Appli
43	74.5	5.9	438	3	US-08-486-099-105	Sequence 105, App
44	74.5	5.9	438	3	US-08-360-107A-115	Sequence 115, App
45	74.5	5.9	438	3	US-08-484-223B-105	Sequence 105, App

ALIGNMENTS

RESULT 1
US-08-485-721-2
; Sequence 2, Application US/08485721
; Patent No. 5821124
; GENERAL INFORMATION:
; APPLICANT: Regeneron Pharmaceuticals, Inc. and
; APPLICANT: Regents of the University of California
; TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,721
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/392,935
; FILING DATE: 02-SEP-1993
; APPLICATION NUMBER: PCT/US93/08326
; FILING DATE: 02-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kempler Ph.D., Gail M.
; REGISTRATION NUMBER: 32,143
; REFERENCE/DOCKET NUMBER: Reg 132
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-347-7000
; TELEFAX: 914-347-2113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-485-721-2

Query Match 100.0%; Score 1268; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 2.3e-126;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERCPSELGTVLVALVVLGRLATPAGGQHYLHIRPAPSDNLPLVDLIEHPDPIFDPKXKD 60
 DB 1 MERCPSELGTVLVALVVLGRLATPAGGQHYLHIRPAPSDNLPLVDLIEHPDPIFDPKXKD 60
 QY 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAGAEADLAELDQLRORPSGAMPSEI 120
 DB 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAGAEADLAELDQLRORPSGAMPSEI 120
 QY 121 KGLEFSEGLAOGKKORLSKRLRKLQMWLMSQTFPCVLYANNDLGSFPMRPYVKGSCFS 180
 DB 121 KGLEFSEGLAOGKKORLSKRLRKLQMWLMSQTFPCVLYANNDLGSFPMRPYVKGSCFS 180
 QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

RESULT 2

US-08-392-935-2
 ; Sequence 2, Application US/08392935
 ; Patent No. 5843775

; GENERAL INFORMATION:

; APPLICANT: Regeneron Pharmaceuticals, Inc. and
 ; APPLICANT: Regents of the University of California
 ; TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
 ; TITLE OF INVENTION: Compositions
 ; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
 ; STREET: 777 Old Saw Mill River Road
 ; CITY: Tarrytown
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10591

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/392,935

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/08326

; FILING DATE: 02-SEP-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Kempler Ph.D., Gail M.

; REGISTRATION NUMBER: 32,143

; REFERENCE/DOCKET NUMBER: Reg 132

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 914-347-7000

; TELEFAX: 914-347-2113

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 232 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-392-935-2

Query Match 100.0%; Score 1268; DB 2; Length 232;

Best Local Similarity 100.0%; Pred. No. 2.3e-126; Indels 0; Gaps 0;

Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERCPSELGTVLVALVVLGRLATPAGGQHYLHIRPAPSDNLPLVDLIEHPDPIFDPKXKD 60
 DB 1 MERCPSELGTVLVALVVLGRLATPAGGQHYLHIRPAPSDNLPLVDLIEHPDPIFDPKXKD 60
 QY 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAGAEADLAELDQLRORPSGAMPSEI 120
 DB 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAGAEADLAELDQLRORPSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKRLRKLQMWLMSQTFPCVLYANNDLGSFPMRPYVKGSCFS 180
 DB 121 KGLEFSEGLAOGKKORLSKRLRKLQMWLMSQTFPCVLYANNDLGSFPMRPYVKGSCFS 180
 QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

RESULT 3

US-08-897-236-2
 ; Sequence 2, Application US/08897236A
 ; Patent No. 6075007

; GENERAL INFORMATION:

; APPLICANT: Regeneron Pharmaceuticals, Inc.
 ; TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition
 ; FILE REFERENCE: REG 133
 ; CURRENT APPLICATION NUMBER: US/08/897,236A
 ; CURRENT FILING DATE: 1997-07-17
 ; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 232

; TYPE: PRT

; ORGANISM: human

US-08-897-236-2

Query Match 100.0%; Score 1268; DB 3; Length 232;

Best Local Similarity 100.0%; Pred. No. 2.3e-126;

Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERCPSELGTVLVALVVLGRLATPAGGQHYLHIRPAPSDNLPLVDLIEHPDPIFDPKXKD 60
 DB 1 MERCPSELGTVLVALVVLGRLATPAGGQHYLHIRPAPSDNLPLVDLIEHPDPIFDPKXKD 60

QY 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAGAEADLAELDQLRORPSGAMPSEI 120
 DB 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAGAEADLAELDQLRORPSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKRLRKLQMWLMSQTFPCVLYANNDLGSFPMRPYVKGSCFS 180
 DB 121 KGLEFSEGLAOGKKORLSKRLRKLQMWLMSQTFPCVLYANNDLGSFPMRPYVKGSCFS 180

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

QY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232
 DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCORRGRCGWIPIQYPIISECKCSC 232

```

; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-167-874-2

Query Match          100.0%; Score 1268; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 2.3e-126;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNLPLVDLIEHDPDPIFDPKKD 60
Db 1 MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNLPLVDLIEHDPDPIFDPKKD 60

QY 61 LNETLLRSLGGHYDPGFMATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120
Db 61 LNETLLRSLGGHYDPGFMATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKKLRKQMWLWSOTFCPVLYAWNDLGSFRFPRYVKVGCFS 180
Db 121 KGLEFSEGLAOGKKORLSKKLRKQMWLWSOTFCPVLYAWNDLGSFRFPRYVKVGCFS 180

QY 181 KRSCSVPEGMVCKPKSKVHLTVLRWRCQRRGQRCGWIPQIYPIIIECKKSC 232
Db 181 KRSCSVPEGMVCKPKSKVHLTVLRWRCQRRGQRCGWIPQIYPIIIECKKSC 232

RESULT 5
PCT-US93-08326-2
; Sequence 2, Application PC/TUS9308326
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
; TITLE OF INVENTION: Compositions
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08326
; FILING DATE: 02-SEP-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kempler Ph.D., Gail M.
; REGISTRATION NUMBER: 32,143
; REFERENCE/DOCKET NUMBER: Reg 132
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-347-7000
; TELEFAX: 914-347-2113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
PCT-US93-08326-2

Query Match          100.0%; Score 1268; DB 5; Length 232;
Best Local Similarity 100.0%; Pred. No. 2.3e-126;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNLPLVDLIEHDPDPIFDPKKD 60
Db 1 MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNLPLVDLIEHDPDPIFDPKKD 60
```

```

QY 61 LNETLLRSLGGHYDPGFMATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120
Db 61 LNETLLRSLGGHYDPGFMATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKKLRKQMWLWSOTFCPVLYAWNDLGSFRFPRYVKVGCFS 180
Db 121 KGLEFSEGLAOGKKORLSKKLRKQMWLWSOTFCPVLYAWNDLGSFRFPRYVKVGCFS 180

QY 181 KRSCSVPEGMVCKPKSKVHLTVLRWRCQRRGQRCGWIPQIYPIIIECKKSC 232
Db 181 KRSCSVPEGMVCKPKSKVHLTVLRWRCQRRGQRCGWIPQIYPIIIECKKSC 232

RESULT 6
US-08-897-236-11
; Sequence 11, Application US/08897236A
; Patent No. 6075007
; GENERAL INFORMATION:
; APPLICANT: Regeneron Pharmaceuticals, Inc.
; TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition
; FILE REFERENCE: REG 133
; CURRENT APPLICATION NUMBER: US/08/897,236A
; CURRENT FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 232
; TYPE: PRT
; ORGANISM: mouse
US-08-897-236-11

Query Match          99.2%; Score 1258; DB 3; Length 232;
Best Local Similarity 99.1%; Pred. No. 2.7e-125;
Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNLPLVDLIEHDPDPIFDPKKD 60
Db 1 MERCPSLGVTLYALVVVLGLRAAPAGGQHYLHIRPAPSDNLPLVDLIEHDPDPIFDPKKD 60

QY 61 LNETLLRSLGGHYDPGFMATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120
Db 61 LNETLLRSLGGHYDPGFMATSPEDRPGGGGAGGAGGAEDLAELDQLLRQPSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKKLRKQMWLWSOTFCPVLYAWNDLGSFRFPRYVKVGCFS 180
Db 121 KGLEFSEGLAOGKKORLSKKLRKQMWLWSOTFCPVLYAWNDLGSFRFPRYVKVGCFS 180

QY 181 KRSCSVPEGMVCKPKSKVHLTVLRWRCQRRGQRCGWIPQIYPIIIECKKSC 232
Db 181 KRSCSVPEGMVCKPKSKVHLTVLRWRCQRRGQRCGWIPQIYPIIIECKKSC 232

RESULT 7
US-09-167-874-11
; Sequence 11, Application US/09167874
; Patent No. 6277593
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al.
; TITLE OF INVENTION: DORSAL TISSUE AFFECTING FACTOR AND COMPOSITIONS
; FILE REFERENCE: REG132-B
; CURRENT APPLICATION NUMBER: US/09/167,874
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/485,721
; EARLIER FILING DATE: 1995-07-06
; EARLIER APPLICATION NUMBER: 08/392,935
; EARLIER FILING DATE: 1995-09-22
; EARLIER APPLICATION NUMBER: PCT/US93/08326
; EARLIER FILING DATE: 1993-09-02
; EARLIER APPLICATION NUMBER: 07/957,401
; EARLIER FILING DATE: 1992-10-06
; EARLIER APPLICATION NUMBER: 07/950,410
; EARLIER FILING DATE: 1992-09-23
; EARLIER APPLICATION NUMBER: 07/939,954
```

EARLIER FILING DATE: 1992-09-03
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 232
TYPE: PRT
ORGANISM: mouse
US-09-167-874-11

Query Match 99.2%; Score 1258; DB 4; Length 232;
Best Local Similarity 99.1%; Pred. No. 2.7e-125;
Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRCPSLGVTLVAVLVVGLRATPAGQHYLHTRPAPSDNLPLVDLIEHPDPIFDPEKXD 60
DB 1 MRCPSLGVTLVAVLVVGLRATPAGQHYLHTRPAPSDNLPLVDLIEHPDPIFDPEKXD 60
QY 61 LNETLRLSLGHIYDGFMAITSPEDRPGGGGAGAEADLAELDQLLRORPSGAMPSEI 120
DB 61 LNETLRLSLGHIYDGFMAITSPEDRPGGGGAGAEADLAELDQLLRORPSGAMPSEI 120
QY 121 KGLEFSEGLAOGKORLSKTLRRKLQWMLWSQTFPCVLYAMNDLGSFPMRYVAVGSCFS 180
DB 121 KGLEFSEGLAOGKORLSKTLRRKLQWMLWSQTFPCVLYAMNDLGSFPMRYVAVGSCFS 180
QY 181 KSCSVPEGVNCCKPSKSVHLTVLRMRCORRGQRCGWIPIQYPIISCKSC 232
DB 181 KSCSVPEGVNCCKPSKSVHLTVLRMRCORRGQRCGWIPIQYPIISCKSC 232

RESULT 8

US-08-897-236-23
Sequence 23, Application US/08897236A
Patent No. 6075007
GENERAL INFORMATION:
APPLICANT: Regeneron Pharmaceuticals, Inc.
TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition
FILE REFERENCE: REG 133
CURRENT APPLICATION NUMBER: US/08/897,236A
CURRENT FILING DATE: 1997-07-17
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 23
LENGTH: 449
TYPE: PRT
ORGANISM: human
US-08-897-236-23

Query Match 94.2%; Score 1195; DB 3; Length 449;
Best Local Similarity 94.8%; Pred. No. 3.1e-118;
Matches 220; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 MRCPSLGVTLVAVLVVGLRATPAGQHYLHTRPAPSDNLPLVDLIEHPDPIFDPEKXD 60
DB 1 MRCPSLGVTLVAVLVVGLRATPAGQHYLHTRPAPSDNLPLVDLIEHPDPIFDPEKXD 60
QY 61 LNETLRLSLGHIYDGFMAITSPEDRPGGGGAGAEADLAELDQLLRORPSGAMPSEI 120
DB 61 LNETLRLSLGHIYDGFMAITSPEDRPGGGGAGAEADLAELDQLLRORPSGAMPSEI 120
QY 121 KGLEFSEGLAOGKORLSKTLRRKLQWMLWSQTFPCVLYAMNDLGSFPMRYVAVGSCFS 180
DB 121 KGLEFSEGLAOGKORLSKTLRRKLQWMLWSQTFPCVLYAMNDLGSFPMRYVAVGSCFS 180
QY 181 KSCSVPEGVNCCKPSKSVHLTVLRMRCORRGQRCGWIPIQYPIISCKSC 232
DB 181 KSCSVPEGVNCCKPSKSVHLTVLRMRCORRGQRCGWIPIQYPIISCKSC 232

RESULT 9
US-08-485-721-9
Sequence 9, Application US/08485721
Patent No. 5821124

GENERAL INFORMATION:
APPLICANT: Regeneron Pharmaceuticals, Inc. and
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,721
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/392,935
FILING DATE: 02-SEP-1993
APPLICATION NUMBER: PCT/US93/08326
FILING DATE: 02-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kempler Ph.D., Gail M.
REGISTRATION NUMBER: 32,143
REFERENCE/DOCKET NUMBER: Reg 132
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-347-7000
TELEFAX: 914-347-2113
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-721-9

Query Match 74.9%; Score 949.5; DB 2; Length 222;
Best Local Similarity 78.1%; Pred. No. 1.2e-92;
Matches 175; Conservative 17; Mismatches 23; Indels 9; Gaps 2;

QY 9 VTLVAVLVVGLRATPAGQHYLHTRPAPSDNLPLVDLIEHPDPIFDPEKXDLNETLRLS 68
DB 8 VTLVAVLVVGLRATPAGQHYLHTRPAPSDNLPLVDLIEHPDPIFDPEKXDLNETLRLS 67
QY 69 LLSGHIDPGFMATSPEDRPGGGGAGAEADLAELDQLLRORPSGAMPSEIKGLEFSEG 128
DB 69 LLSGHIDPGFMATSPEDRPGGGGAGAEADLAELDQLLRORPSGAMPSEIKGLEFSEG 119
QY 129 LAGCKORLSKTLRRKLQWMLWSQTFPCVLYAMNDLGSFPMRYVAVGSCFSKSCSYPE 188
DB 120 L-OSKGRHSKTLRRKLQWMLWSQTFPCVLYAMNDLGSFPMRYVAVGSCFSKSCSYPE 178
QY 189 GNVCKPSKSVHLTVLRMRCORRGQRCGWIPIQYPIISCKSC 232
DB 179 GNVCKPSKSVHLTVLRMRCORRGQRCGWIPIQYPIISCKSC 222

RESULT 10
US-08-392-935-9
Sequence 9, Application US/08392935
Patent No. 5843775
GENERAL INFORMATION:
APPLICANT: Regeneron Pharmaceuticals, Inc. and
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
NUMBER OF SEQUENCES: 24

ATTORNEY/AGENT INFORMATION:
NAME: Kempster Ph.D., Gail M.
REGISTRATION NUMBER: 32,143
REFERENCE/DOCKET NUMBER: Reg 132
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-347-7000
TELEFAX: 914-347-2113
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 87 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-721-11

Query Match 39.7%; Score 503; DB 2; Length 87;
Best Local Similarity 98.9%; Pred. No. 5.9e-46;
Matches 86; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 146 QMMLMSQTFPCPVLYANNDLSRFPWRYVYKSCFSKRSQSVPEGWCKPSKSVSLTVLRW 205
Db 1 QMMLMSQTFPCPVLYANNDLSRFPWRYVYKSCFSKRSQSVPEGWCKPSKSVSLTVLRW 60
Qy 206 RCORRGQRCGWIPIOYPIISECKCSC 232
Db 61 RCORRGQRCGWIPIOYPIISECKCSC 87

RESULT 13
US-08-392-935-11
Sequence 11, Application US/08392935
Patent No. 5843775
GENERAL INFORMATION:
APPLICANT: Regeneron Pharmaceuticals, Inc. and
APPLICANT: Regents of the University of California
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
TITLE OF INVENTION: Compositions
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,935
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08326
FILING DATE: 02-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kempster Ph.D., Gail M.
REGISTRATION NUMBER: 32,143
REFERENCE/DOCKET NUMBER: Reg 132
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-347-7000
TELEFAX: 914-347-2113
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 87 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-392-935-11

Query Match 39.7%; Score 503; DB 2; Length 87;

Best Local Similarity 98.9%; Pred. No. 5.9e-46;
Matches 86; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 146 QMMLMSQTFPCPVLYANNDLSRFPWRYVYKSCFSKRSQSVPEGWCKPSKSVSLTVLRW 205
Db 1 QMMLMSQTFPCPVLYANNDLSRFPWRYVYKSCFSKRSQSVPEGWCKPSKSVSLTVLRW 60
Qy 206 RCORRGQRCGWIPIOYPIISECKCSC 232
Db 61 RCORRGQRCGWIPIOYPIISECKCSC 87

RESULT 14
PCT-US93-08326-11
Sequence 11, Application PC/TUS9308326
GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and
TITLE OF INVENTION: Compositions
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08326
FILING DATE: 02-SEP-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kempster Ph.D., Gail M.
REGISTRATION NUMBER: 32,143
REFERENCE/DOCKET NUMBER: Reg 132
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-347-7000
TELEFAX: 914-347-2113
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 87 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US93-08326-11

Query Match 39.7%; Score 503; DB 5; Length 87;
Best Local Similarity 98.9%; Pred. No. 5.9e-46;
Matches 86; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 146 QMMLMSQTFPCPVLYANNDLSRFPWRYVYKSCFSKRSQSVPEGWCKPSKSVSLTVLRW 205
Db 1 QMMLMSQTFPCPVLYANNDLSRFPWRYVYKSCFSKRSQSVPEGWCKPSKSVSLTVLRW 60
Qy 206 RCORRGQRCGWIPIOYPIISECKCSC 232
Db 61 RCORRGQRCGWIPIOYPIISECKCSC 87

RESULT 15
US-08-586-165-5
Sequence 5, Application US/08586165
Patent No. 6054298
GENERAL INFORMATION:
APPLICANT: Laufer, Edward M.
APPLICANT: Orozco, Olivia E.
APPLICANT: Tabin, Clifford J.
TITLE OF INVENTION: Fringe Proteins and Pattern Formation

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 09:24:43 ; Search time 10 Seconds
(without alignments)
439.675 Million cell updates/sec

Title: US-09-897-322-2
Perfect score: 1268
Sequence: 1 MERCPSLGVTLYALVVVLGL.....QRCGWIPQIPIISECKCSC 232

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues
Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/1/pubpaa/US08_NEW PUB.pdb.p*
- 2: /cgn2_6/prodata/1/pubpaa/ECT_NEW PUB.pdb.p*
- 3: /cgn2_6/prodata/1/pubpaa/US06_NEW PUB.pdb.p*
- 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pdb.p*
- 5: /cgn2_6/prodata/1/pubpaa/US07_NEW PUB.pdb.p*
- 6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pdb.p*
- 7: /cgn2_6/prodata/1/pubpaa/ECTUS_PUBCOMB.pdb.p*
- 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pdb.p*
- 9: /cgn2_6/prodata/1/pubpaa/US09_NEW PUB.pdb.p*
- 10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pdb.p*
- 11: /cgn2_6/prodata/1/pubpaa/US10_NEW PUB.pdb.p*
- 12: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pdb.p*
- 13: /cgn2_6/prodata/1/pubpaa/US60_NEW PUB.pdb.p*
- 14: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pdb.p*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1268	100.0	232	9	US-10-044-716-4
2	1268	100.0	232	10	US-09-897-322-2
3	1258	99.2	232	9	US-10-044-716-6
4	1258	99.2	232	10	US-09-897-322-11
5	90.5	7.1	623	10	US-09-854-731-4
6	88.5	7.0	979	9	US-09-884-001-4
7	87	6.9	14	10	US-09-897-322-3
8	86.5	6.8	939	10	US-09-854-845-16
9	86.5	6.8	954	10	US-09-854-845-14
10	86.5	6.8	1034	10	US-09-854-845-6
11	86.5	6.8	1049	10	US-09-854-845-2
12	86.5	6.8	1078	10	US-09-854-845-8
13	86.5	6.8	1093	10	US-09-854-845-4
14	86.5	6.8	1136	10	US-09-854-845-12
15	86.5	6.8	1151	10	US-09-854-845-10
16	86	6.8	625	10	US-09-854-731-18
17	83	6.5	538	10	US-09-827-822-9
18	80.5	6.3	465	9	US-10-029-180-44
19	80.5	6.3	938	9	US-10-029-217A-28

20	80.5	6.3	938	9	US-10-029-217A-29	Sequence 29, Appl
21	79	6.2	264	9	US-09-738-626-5623	Sequence 5623, Ap
22	79	6.2	325	9	US-10-063-547-64	Sequence 64, Appl
23	79	6.2	325	10	US-09-741-711-2	Sequence 2, Appl
24	79	6.2	325	12	US-10-006-867-64	Sequence 64, Appl
25	79	6.2	325	12	US-10-052-586-238	Sequence 238, App
26	79	6.2	11877	9	US-09-860-846-6	Sequence 6, Appl
27	79	6.2	11877	10	US-09-861-289-6	Sequence 6, Appl
28	78	6.1	135	10	US-09-925-297-852	Sequence 852, App
29	77	6.1	428	9	US-09-734-329-2	Sequence 2, Appl
30	77	6.1	1249	9	US-09-964-899-33	Sequence 33, Appl
31	76.5	6.0	144	10	US-09-811-384-131	Sequence 131, App
32	76	6.0	14	10	US-09-897-322-5	Sequence 5, Appl
33	76	6.0	169	10	US-09-823-038A-53	Sequence 53, Appl
34	76	6.0	588	10	US-09-782-980-78	Sequence 78, Appl
35	74.5	5.9	321	10	US-09-925-301-938	Sequence 938, App
36	74.5	5.9	550	10	US-09-873-233A-20	Sequence 20, Appl
37	74	5.8	848	9	US-09-882-986-2	Sequence 2, Appl
38	74	5.8	4613	9	US-09-860-846-31	Sequence 31, Appl
39	74	5.8	4613	10	US-09-861-289-31	Sequence 31, Appl
40	73.5	5.8	127	10	US-09-795-006A-47	Sequence 47, Appl
41	73.5	5.8	295	9	US-09-978-295A-2	Sequence 2, Appl
42	73.5	5.8	295	9	US-09-978-697-2	Sequence 2, Appl
43	73.5	5.8	295	9	US-09-978-132A-2	Sequence 2, Appl
44	73.5	5.8	295	9	US-09-999-832A-2	Sequence 2, Appl
45	73	5.8	12	10	US-09-897-322-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-10-044-716-4
; Sequence 4, Application US/10044716
; Patent No. US20020159986A1
; GENERAL INFORMATION:
; APPLICANT: LANGENFELD, John
; FILE OF INVENTION: BONE MORPHOGENETIC PROTEIN-2 IN THE TREATMENT AND DIAGNOSIS OF CA
; TITLE REFERENCE: 270/070US
; CURRENT APPLICATION NUMBER: US/10/044,716
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US60/261,252
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-044-716-4

Query Match	100.0%	Score 1268;	DB 9;	Length 232;
Best Local Similarity	100.0%;	Pred. No. 2.8e-108;	Mismatches 0;	Indels 0;
Matches 232;	Conservative	0;	Gaps 0;	
Qy	1	MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDPIFDPKED	60	
Db	1	MERCPSLGVTLYALVVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDPIFDPKED	60	
Qy	61	LNETLRLSLGGHYDPGFWMATSPEDRPGCGGGAAGGAEDLAELDOLLQRPSPGAMPSEI	120	
Db	61	LNETLRLSLGGHYDPGFWMATSPEDRPGCGGGAAGGAEDLAELDOLLQRPSPGAMPSEI	120	
Qy	121	KGLEFSEGLAOCKKQKLSKLRKLMWLSOTFCPLVYAWNDLGSRRFWRPVYKVGSCFS	180	
Db	121	KGLEFSEGLAOCKKQKLSKLRKLMWLSOTFCPLVYAWNDLGSRRFWRPVYKVGSCFS	180	
Qy	181	KRSCSVPEGMVCKPSKSVHLTVLRWRCQRGGRCRCWIPQIPIIIECKCSC	232	
Db	181	KRSCSVPEGMVCKPSKSVHLTVLRWRCQRGGRCRCWIPQIPIIIECKCSC	232	
RESULT 2				

```
US-09-897-322-2
; Sequence 2, Application US/09897322
; Patent No. US20020102643A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al.
; TITLE OF INVENTION: DORSAL TISSUE AFFECTING FACTOR AND COMPOSITIONS
; FILE REFERENCE: REG132-B
; CURRENT APPLICATION NUMBER: US/09/897,322
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/167,874
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 08/485,721
; PRIOR FILING DATE: 1995-07-06
; PRIOR APPLICATION NUMBER: 08/392,935
; PRIOR FILING DATE: 1995-09-22
; PRIOR APPLICATION NUMBER: PCT/US93/08326
; PRIOR FILING DATE: 1993-09-02
; PRIOR APPLICATION NUMBER: 07/957,401
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: 07/950,410
; PRIOR FILING DATE: 1992-09-23
; PRIOR APPLICATION NUMBER: 07/939,954
; PRIOR FILING DATE: 1992-09-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-322-2
```

```
Query Match      100.0%; Score 1268; DB 10; Length 232;
Best Local Similarity 100.0%; Pred. No. 2,3e-108;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYLVVVLGIRATPAGGQHYLHIRPAPSDNPLVLDLIEHPDIPFPKED 60
Db 1 MERCPSLGVTLYLVVVLGIRATPAGGQHYLHIRPAPSDNPLVLDLIEHPDIPFPKED 60

QY 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAAGAEADLAEIDQLLRORSGAMPSEI 120
Db 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAAGAEADLAEIDQLLRORSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKLRKLQMWLMSOTFCFVLVYAMNDLSGRFWRVYKVGSCFS 180
Db 121 KGLEFSEGLAOGKKORLSKLRKLQMWLMSOTFCFVLVYAMNDLSGRFWRVYKVGSCFS 180

QY 181 KRCSVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPFIQYPIISECKCSC 232
Db 181 KRCSVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPFIQYPIISECKCSC 232
```

```
RESULT 3
US-10-044-716-6
; Sequence 6, Application US/10044716
; Patent No. US2002015986A1
; GENERAL INFORMATION:
; APPLICANT: LANGENFELD, John
; TITLE OF INVENTION: BONE MORPHOGENETIC PROTEIN-2 IN THE TREATMENT AND DIAGNOSIS OF CA
; FILE REFERENCE: 270/070US
; CURRENT APPLICATION NUMBER: US/10/044,716
; PRIOR FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US60/261,252
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-044-716-6
```

Query Match 99.2%; Score 1258; DB 9; Length 232;

```
Best Local Similarity 99.1%; Pred. No. 2,3e-107;
Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYLVVVLGIRATPAGGQHYLHIRPAPSDNPLVLDLIEHPDIPFPKED 60
Db 1 MERCPSLGVTLYLVVVLGIRAPAGGQHYLHIRPAPSDNPLVLDLIEHPDIPFPKED 60

QY 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAAGAEADLAEIDQLLRORSGAMPSEI 120
Db 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAEADLAEIDQLLRORSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKLRKLQMWLMSOTFCFVLVYAMNDLSGRFWRVYKVGSCFS 180
Db 121 KGLEFSEGLAOGKKORLSKLRKLQMWLMSOTFCFVLVYAMNDLSGRFWRVYKVGSCFS 180

QY 181 KRCSVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPFIQYPIISECKCSC 232
Db 181 KRCSVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPFIQYPIISECKCSC 232
```

```
RESULT 4
US-09-897-322-11
; Sequence 11, Application US/09897322
; Patent No. US20020102643A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al.
; TITLE OF INVENTION: DORSAL TISSUE AFFECTING FACTOR AND COMPOSITIONS
; FILE REFERENCE: REG132-B
; CURRENT APPLICATION NUMBER: US/09/897,322
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/167,874
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 08/485,721
; PRIOR FILING DATE: 1995-07-06
; PRIOR APPLICATION NUMBER: 08/392,935
; PRIOR FILING DATE: 1995-09-22
; PRIOR APPLICATION NUMBER: PCT/US93/08326
; PRIOR FILING DATE: 1993-09-02
; PRIOR APPLICATION NUMBER: 07/957,401
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: 07/950,410
; PRIOR FILING DATE: 1992-09-23
; PRIOR APPLICATION NUMBER: 07/939,954
; PRIOR FILING DATE: 1992-09-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 232
; TYPE: PRT
; ORGANISM: mouse
US-09-897-322-11
```

```
Query Match      99.2%; Score 1258; DB 10; Length 232;
Best Local Similarity 99.1%; Pred. No. 2,3e-107;
Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYLVVVLGIRATPAGGQHYLHIRPAPSDNPLVLDLIEHPDIPFPKED 60
Db 1 MERCPSLGVTLYLVVVLGIRAPAGGQHYLHIRPAPSDNPLVLDLIEHPDIPFPKED 60

QY 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAAGAEADLAEIDQLLRORSGAMPSEI 120
Db 61 LNETLLRSLLGGHYDPGFMAATSPEDRPGGGGAGAEADLAEIDQLLRORSGAMPSEI 120

QY 121 KGLEFSEGLAOGKKORLSKLRKLQMWLMSOTFCFVLVYAMNDLSGRFWRVYKVGSCFS 180
Db 121 KGLEFSEGLAOGKKORLSKLRKLQMWLMSOTFCFVLVYAMNDLSGRFWRVYKVGSCFS 180

QY 181 KRCSVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPFIQYPIISECKCSC 232
Db 181 KRCSVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPFIQYPIISECKCSC 232
```

Db 685 TPTTTPHSGD-----ICTSWPRPIFGSLHHVDPDLSCRGHHTTILIVKVLNKSXTIRSNSSG 733
 QY 73 -HYDPGFMTATPPEDRPGGGGAAGAEADLAELDQLLRQPSGAMPSEIKGLEFSEGL-- 129
 Db 740 LSIQTVFQSS-----PGGGGGGGGEEEDSQSETPD-PSGFRGTWEADRGWGLIS 793
 QY 130 ---AQQKQRLSKLRRKLOMMLWSQTFPCVLYAYNNDLGSRF 168
 Db 794 PTEAMGNSNGASSSCFGWLRKELENAEFTMPDPSPLSAAF 835
 RESULT 7
 US-09-897-322-3
 ; Sequence 3, Application US/09897322
 ; Patent No. US20020102643A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al.
 ; TITLE OF INVENTION: DORSAL TISSUE AFFECTING FACTOR AND COMPOSITIONS
 ; FILE REFERENCE: REG132-B
 ; CURRENT APPLICATION NUMBER: US/09/897,322
 ; CURRENT FILING DATE: 2001-07-02
 ; PRIOR APPLICATION NUMBER: 09/167,874
 ; PRIOR FILING DATE: 1998-10-07
 ; PRIOR APPLICATION NUMBER: 08/485,721
 ; PRIOR FILING DATE: 1995-07-06
 ; PRIOR APPLICATION NUMBER: 08/392,935
 ; PRIOR FILING DATE: 1995-09-22
 ; PRIOR APPLICATION NUMBER: PCT/US93/083326
 ; PRIOR FILING DATE: 1993-09-02
 ; PRIOR APPLICATION NUMBER: 07/957,401
 ; PRIOR FILING DATE: 1992-10-06
 ; PRIOR APPLICATION NUMBER: 07/950,410
 ; PRIOR FILING DATE: 1992-09-23
 ; PRIOR APPLICATION NUMBER: 07/939,954
 ; PRIOR FILING DATE: 1992-09-03
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 14
 ; TYPE: PRT
 ; ORGANISM: frog and mouse
 US-09-897-322-3
 Query Match 6.9%; Score 87; DB 10; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0
 QY 146 QMWLWSQTFPCVLY 159
 Db 1 QMWLWSQTFPCVLY 14
 RESULT 8
 US-09-854-845-16
 ; Sequence 16, Application US/09854845
 ; Patent No. US20020098491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Wang, Xiaoming
 ; APPLICANT: Scoville, John
 ; APPLICANT: Turner, C. Alexander Jr.
 ; TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and
 ; FILE REFERENCE: LEX-0177-USA
 ; CURRENT APPLICATION NUMBER: US/09/854,845
 ; CURRENT FILING DATE: 2001-05-14
 ; PRIOR APPLICATION NUMBER: US 60/205,274
 ; PRIOR FILING DATE: 2000-05-18
 ; PRIOR APPLICATION NUMBER: US 60/208,893
 ; PRIOR FILING DATE: 2000-06-02
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 16
 ; LENGTH: 939

```
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-854-845-16
Query Match
Best Local Similarity 30.2%; Pred. No. 5.5;
Matches 35; Conservative 10; Mismatches 44; Indels 27; Gaps 9;

Oy 132 GKQRLSKLRLQWMLWSQ--TFCPVLYAMNDLGSFMPRYV-----KVGSCFSK 181
Db 380 GKQQRCS-TLEDSSNMSLWNTACPVNRVTRDGFSGFSPWQPCHELDGNSGCLCR 438

Oy 182 -RSCSVPE---GMVCKPSKSVHLTVLRMRCQRRGQRCGMPIQYPIISECKCSC 232
Db 439 ARSCDSPPRCGLDCL-GPAIHIA---NCSRNG---AWTP--WSSWALCSTSC 483

RESULT 9
US-09-854-845-14
; Sequence 14, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wang, Xiaoming
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020098491A1 Human Semaphorin Homologs and Polynucleotid
; FILE REFERENCE: LEX-0177-USA
; CURRENT APPLICATION NUMBER: US/09/854,845
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/205,274
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 60/208,893
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 954
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-854-845-14
Query Match
Best Local Similarity 30.2%; Pred. No. 5.6; DB 10; Length 954;
Matches 35; Conservative 10; Mismatches 44; Indels 27; Gaps 9;

Oy 132 GKQRLSKLRLQWMLWSQ--TFCPVLYAMNDLGSFMPRYV-----KVGSCFSK 181
Db 380 GKQQRCS-TLEDSSNMSLWNTACPVNRVTRDGFSGFSPWQPCHELDGNSGCLCR 438

Oy 182 -RSCSVPE---GMVCKPSKSVHLTVLRMRCQRRGQRCGMPIQYPIISECKCSC 232
Db 439 ARSCDSPPRCGLDCL-GPAIHIA---NCSRNG---AWTP--WSSWALCSTSC 483

RESULT 10
US-09-854-845-6
; Sequence 6, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wang, Xiaoming
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020098491A1 Human Semaphorin Homologs and Polynucleotid
; FILE REFERENCE: LEX-0177-USA
; CURRENT APPLICATION NUMBER: US/09/854,845
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/205,274
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 60/208,893
; NUMBER OF SEQ ID NOS: 50
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-854-845-6
Query Match
Best Local Similarity 30.2%; Pred. No. 6.2; DB 10; Length 1034;
Matches 35; Conservative 10; Mismatches 44; Indels 27; Gaps 9;

Oy 132 GKQRLSKLRLQWMLWSQ--TFCPVLYAMNDLGSFMPRYV-----KVGSCFSK 181
Db 475 GKQQRCS-TLEDSSNMSLWNTACPVNRVTRDGFSGFSPWQPCHELDGNSGCLCR 533

Oy 182 -RSCSVPE---GMVCKPSKSVHLTVLRMRCQRRGQRCGMPIQYPIISECKCSC 232
Db 534 ARSCDSPPRCGLDCL-GPAIHIA---NCSRNG---AWTP--WSSWALCSTSC 578

RESULT 11
US-09-854-845-2
; Sequence 2, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wang, Xiaoming
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020098491A1 Human Semaphorin Homologs and Polynucleotid
; FILE REFERENCE: LEX-0177-USA
; CURRENT APPLICATION NUMBER: US/09/854,845
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/205,274
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 60/208,893
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-854-845-2
Query Match
Best Local Similarity 30.2%; Pred. No. 6.3; DB 10; Length 1049;
Matches 35; Conservative 10; Mismatches 44; Indels 27; Gaps 9;

Oy 132 GKQRLSKLRLQWMLWSQ--TFCPVLYAMNDLGSFMPRYV-----KVGSCFSK 181
Db 475 GKQQRCS-TLEDSSNMSLWNTACPVNRVTRDGFSGFSPWQPCHELDGNSGCLCR 533

Oy 182 -RSCSVPE---GMVCKPSKSVHLTVLRMRCQRRGQRCGMPIQYPIISECKCSC 232
Db 534 ARSCDSPPRCGLDCL-GPAIHIA---NCSRNG---AWTP--WSSWALCSTSC 578

RESULT 12
US-09-854-845-8
; Sequence 8, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wang, Xiaoming
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020098491A1 Human Semaphorin Homologs and Polynucleotid
; FILE REFERENCE: LEX-0177-USA
; CURRENT APPLICATION NUMBER: US/09/854,845
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/205,274
; PRIOR FILING DATE: 2000-05-18
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 09:22:07 ; Search time 33 Seconds

(without alignments)
1448.574 Million cell updates/sec

Title: US-09-897-322-2

Perfect score: 1268

Sequence: 1 MERCPSLGVTLYALVVVLGL.....QRCGWIPQYPIISECKKSC 232

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	102	8.0	2393	2 Q9Z5F4	Q9Z5F4 myxococcus
2	99.5	7.8	413	4 Q9UBU6	Q9UBU6 homo sapien
3	96	7.6	507	10 Q93VF7	Q93VF7 arabidopsis
4	94.5	7.5	684	2 Q44055	Q44055 aeromonas h
5	93.5	7.4	322	16 Q9RK51	Q9RK51 streptomyce
6	93.5	7.4	372	13 Q12970	Q12970 gallus gall
7	93.5	7.4	372	13 Q12972	Q12972 gallus gall
8	92	7.3	538	15 Q9WJP4	Q9WJP4 moloney mur
9	92	7.3	1737	15 Q92808	Q92808 moloney mur
10	90	7.1	142	10 Q42322	Q42322 arabidopsis
11	88.5	7.0	931	12 Q9QAP6	Q9QAP6 rangiferine
12	88.5	7.0	979	4 Q8TD19	Q8TD19 homo sapien
13	88.5	7.0	979	4 Q8TCY4	Q8TCY4 homo sapien
14	88	6.9	278	16 Q8UGW7	Q8UGW7 agrobacteri
15	88	6.9	814	2 Q59129	Q59129 arthrobacte
16	88	6.9	816	2 Q93NH5	Q93NH5 arthrobacte

17	87.5	6.9	521	2 Q9KI55	Q9KI55 pseudomonas
18	87.5	6.9	728	4 Q9UPH9	Q9UPH9 homo sapien
19	87.5	6.9	1386	5 Q9NKJ2	Q9NKJ2 leishmania
20	87	6.9	1975	5 Q9VCD1	Q9VCD1 drosophila
21	86.5	6.8	351	10 Q94GL5	Q94GL5 oryza sativ
22	86.5	6.8	420	10 Q9ZSY5	Q9ZSY5 arabidopsis
23	86.5	6.8	457	16 Q9RRJ4	Q9RRJ4 deinococcus
24	86.5	6.8	529	2 P94909	P94909 microbacter
25	86.5	6.8	622	10 Q9ZSY6	Q9ZSY6 arabidopsis
26	86.5	6.8	622	10 Q9C6B8	Q9C6B8 arabidopsis
27	86.5	6.8	1202	4 Q9P283	Q9P283 homo sapien
28	86	6.8	525	10 P93520	P93520 zea mays (m
29	86	6.8	981	15 Q92809	Q92809 abelson mur
30	85	6.7	599	10 Q41791	Q41791 zea mays (m
31	85	6.7	1621	4 Q13395	Q13395 homo sapien
32	85	6.7	5146	6 Q8SPM4	Q8SPM4 bos taurus
33	84.5	6.7	212	13 Q9W7C5	Q9W7C5 gallus gall
34	84	6.6	363	17 Q9YAX6	Q9YAX6 aeropyrum p
35	84	6.6	442	13 Q9DFG7	Q9DFG7 brachydanio
36	84	6.6	640	10 Q96397	Q96397 chlamydomon
37	83	6.5	266	5 Q18862	Q18862 caenorhabdi
38	83	6.5	476	12 Q80890	Q80890 herpesvirus
39	83	6.5	538	15 Q56222	Q56222 moloney mur
40	83	6.5	1066	10 Q22804	Q22804 arabidopsis
41	83	6.5	2111	3 Q9C102	Q9C102 schizosacch
42	82	6.5	122	16 Q9RSP2	Q9RSP2 deinococcus
43	82	6.5	339	4 Q00320	Q00320 homo sapien
44	82	6.5	344	16 Q9RZM1	Q9RZM1 deinococcus
45	82	6.5	368	13 Q90853	Q90853 gallus gall

ALIGNMENTS

RESULT 1

Q9Z5F4	PRELIMINARY;	PRT; 2393 AA.
ID	Q9Z5F4	
AC	Q9Z5F4;	
DT	01-MAY-1999 (Tremblrel. 10, Created)	
DT	01-MAY-1999 (Tremblrel. 10, Last sequence update)	
DT	01-MAY-2002 (Tremblrel. 20, Last annotation update)	
DE	Tal (Fragment).	
GN	TAL.	
OS	Myxococcus xanthus.	
OC	Bacteria; Proteobacteria; delta subdivision; Myxobacteria;	
OC	Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.	
OX	NCBI_TaxID=34;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ER-15;	
RX	MEDLINE=99141200; PubMed=9973564;	
RA	Paitan Y., Alon G., Orr E., Ron E.Z., Rosenberg E.;	
RT	"The first gene in the biosynthesis of the polyketide antibiotic TA of	
RT	Myxococcus xanthus codes for a unique PKS module coupled to a peptide	
RT	synthetase."	
RL	J. Mol. Biol. 286:465-474(1999).	
DR	EMBL; AJ006977; CAB38084.1; .	
DR	HSSP; P14687; 1AMU.	
DR	InterPro; IPR000873; AMP-bind.	
DR	InterPro; IPR001242; Condensatn.	
DR	InterPro; IPR000794; Ketoacyl-synt.	
DR	InterPro; IPR003880; Ppanthe attach.	
DR	Pfam; PF00501; AMP-binding; 1.	
DR	Pfam; PF00668; Condensation; 1.	
DR	Pfam; PF00109; ketoacyl-synt; 2.	
DR	Pfam; PF02801; ketoacyl-synt_C; 1.	
DR	Pfam; PF00550; pp-binding; 2.	
DR	PRINTS; PR00154; AMPBINDING.	
DR	PROSITE; PS00075; ACP DOMAIN; 2.	
DR	PROSITE; PS00455; AMP_BINDING; 1.	
FT	NON_TER 1	
FT	NON_TER 2393	
SQ	SEQUENCE 2393 AA; 259537 MW; 6CF6E0153BC3E126 CRC64;	

Query Match 8.0%; Score 102; DB 2; Length 2393;
 Best Local Similarity 23.7%; Pred. No. 1.6;
 Matches 56; Conservative 27; Mismatches 67; Indels 86; Gaps 12;

7 LGVTALVVLGLRATPAGQH-----YLHTRPAPSDNPLVLDLIEHPDIFDEREK 59
 DB LATAICGFNVMAVARAGL-GQHSFGLRLH-----DS--VIDGLEHMYFFPRVVK 294

QY 60 DINETLRLSLGGHYDPG-----MATSPEDRPGGGGAGAEADLADOLLR 109
 DB 295 D-----LRLSNGPEEAPGFQMTFTFOSLQLTSAFPRPEPSGG-----LPLEPLDC 341

QY 110 QPSSGAMPEIKGLFSEGLAQGK-----ORLSKLRKLQW----- 147
 DB 342 VHOEGAYPLELEVVEGKGLTLHKYDARLYEADTVERRMRQLRAADQYADVESPLA 401

QY 148 --WL-----WSQFCPVLYAMNDLSRFWPRYVKGSCFSKSGSVPEGM 190
 DB 402 LSWLDEERRTLLRDMNATATPPL---EDLG-----VHELFGQARETPDAM 445

RESULT 2
 QY 09UB6 PRELIMINARY; PRT; 413 AA.
 AC QY 09UB6;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE Autosomal highly conserved protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bourgeon T., Jamin S., Leroy P., McElreavey K., Fellous M.;
 RT "Identification of AACP, a novel highly conserved gene on human
 RT chromosome 6 candidate for genetic predisposition to schizophrenia.",
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF097027; AAF07850.1; -;
 DR EMBL; AF097026; AAF07849.1; -;
 SQ SEQUENCE 413 AA; 44123 MW; 35B92497907CCD2E CRC64;

Query Match 7.8%; Score 99.5; DB 4; Length 413;
 Best Local Similarity 27.6%; Pred. No. 0.32;
 Matches 50; Conservative 17; Mismatches 69; Indels 45; Gaps 11;

QY 26 GGOHYHTRPAPSDNPLVLDLIEHP--DPIFDEREKDNLRLSLGGHYDPGFMATSP 83
 DB 19 GGDH-----BPVSLRGPPTTAVPCPRDDPAEPPQA-----PG----- 51

QY 84 PDRPGGGGAGAEADLADOLLR-----PSGAMPEIKGLFSEGLA--QGKQRL 137
 DB 52 ---RPAPELAAAAADKLEPRELKRKGAAGSGAELBOQACAPALAPRERARL 108

QY 138 S-KLRLKRLQWLMWSQFCPVLYAMNDLSRFWPRYV--VSCFSKSGSVPEGNVCK 193
 DB 109 SAREYSRQVHEWLM--QSYCYL--TWHS--GLAAPPAYCSPSPQSPFGGAAVQAAPP 165

QY 194 P 194
 DB 166 P 166

RESULT 3
 QY 093VF7 PRELIMINARY; PRT; 507 AA.
 AC QY 093VF7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Putative WD-40 repeat protein, MS14 (At2g19520/F3P11.12).
 GN AT2G19520 OR F3P11.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Euphorbia II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
 RA Sakurai T., Theologis A., Davis R.W.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.W., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Tortum M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL; AY059799; AAL24281.1; -;
 DR EMBL; AY057655; AAL15286.1; -;
 DR EMBL; AY081447; AAM10009.1; -;
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 6.
 DR ProDom; PD0000018; WD40; 1.
 DR PROSITE; PS00882; WD_REPEATS_2; 3.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 DR Repeat; WD repeat.
 SQ SEQUENCE 507 AA; 55759 MW; C37F8000F8B33397 CRC64;

Query Match 7.6%; Score 96; DB 10; Length 507;
 Best Local Similarity 30.5%; Pred. No. 0.89;
 Matches 29; Conservative 10; Mismatches 40; Indels 16; Gaps 4;

QY 80 ATSPEDRPGGGGAG-----GAEDLADOLLRPSGAMPEIKGLFSEGLAQGKKQ 135
 DB 8 AVSPQATTSSGCGAGSPKRGKRPKTKEDS---QTPSSQGSQSVKXME-----SGKTT 58

QY 136 RLKSLKRLKRLQWLMWSQFCPVLYAMNDLSRFWPR 170
 DB 59 QQSPSVDEKYSQW---KGLVPLLYDMLANHLNLPW 90

RESULT 4
 QY 044055 PRELIMINARY; PRT; 684 AA.
 AC Q44055;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Phospholipase C.
 OS Aeromonas hydrophila.
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.


```

OX NCBI_TaxID=644;
RN [1]
RP SEQUENCE OF 1-619 FROM N.A.
RC STRAIN=JMP636;
RX MEDLINE=95284718; PubMed=7767226;
RA Ingham A.B.; Pemberton J.M.;
RT "A lipase of Aeromonas hydrophila showing nonhemolytic phospholipase C
activity.";
RL Curr. Microbiol. 31:28-33(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JMP636;
RA Ingham A.B.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U14011; AAA75598.1; -.
DR InterPro: IPR000734; Lipase.
DR PROSITE; PS00120; LIPASE_SER; 1.
SQ SEQUENCE 684 AA; 72396 MW; 702745900A9A5714 CRC64;

Query Match 7.5%; Score 94.5; DB 2; Length 684;
Best Local Similarity 25.2%; Pred. No. 1.8;
Matches 58; Conservative 22; Mismatches 89; Indels 61; Gaps 10;

QY 19 GLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDIPDPKEKDLNETLLSLGG---HYD 75
Db 289 GSRVAPASGNKDA-IATVAPSRPLAELRPRFIPGRSSCLTSSPRLAPLVPGIKRPS 347
QY 76 PGFMATSP-----PEDRPGG---GGG-----AAGGAEDLAELQ 106
Db 348 PGTVPSQPVHRQCAAGVTVDRRPGRRCGSLANGTDCRSDSPRAAGGQEAHRGDP 407
QY 107 LLRQRPSGAMPSEIKLEFSEGLAQKKORLSKKLRKQLMWLWSOTFCPLVYAMNDL-- 164
Db 408 HLGRQAGSR-AEYRPLQPAAGRAIRADAHLCRRRAV-----CHHRCDLHLSAHHLGE 462
QY 165 GSRFWPRYKVGSCFSKRSKSPVEGVMCKPSKVHLTVLRWCQRGGQR 214
Db 463 GERLCP-----GAGPDRCRCSKHGQEC-----RRGGDR 491

RESULT 5
Q9RK51 PRELIMINARY; PRT; 322 AA.
ID Q9RK51
AC Q9RK51;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE Hypothetical protein SC00326.
GN SC00326 OR SCF12.05.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D.; Chater K.F.; Cerdano-Tarraga A.-M.; Challis G.L.;
RA Thomson N.R.; James K.D.; Harris D.E.; Quail M.A.; Kieser H.;
RA Harper D.; Bateman A.; Brown S.; Chandra G.; Chen C.W.; Collins M.;
RA Cronin A.; Fraser A.; Goble A.; Hidalgo J.; Hornsby T.; Howarth S.;
RA Huang C.-H.; Kieser T.; Larke L.; Murphy L.; Oliver K.; O'Neill S.;
RA Rabinowitz E.; Rajandream M.A.; Rutherford K.; Rutter S.;
RA Seeger K.; Saunders D.; Sharp S.; Squares S.; Taylor K.;
RA Warren T.; Wietzorrek A.; Woodward J.; Barrell B.G.; Parkhill J.;
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL117669; CAB56131.1; -.
KW Hypothetical protein.
SQ SEQUENCE 322 AA; 35339 MW; DD55BB0480090638 CRC64;

Query Match 7.4%; Score 93.5; DB 16; Length 322;

```

```

Best Local Similarity 26.3%; Pred. No. 0.88;
Matches 45; Conservative 21; Mismatches 50; Indels 55; Gaps 10;

QY 9 VTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDIPDP--KEKDLNETLL 66
Db 131 LAVHVLVTAAGVVRTP-----HFVGAEPARTGN-----GPRGLRPLGPEDLARELI 178
QY 67 RSLGGHYDPGFMATSPPEDRPGGGGAAGGAEDLAELDQLLRQRP---SGAMPSEIKGL 123
Db 179 TSLGGTARASALFAATPPD-----DILTRADFPADPGLLP---EGL 216
QY 124 EFSEGLAQKKORLSKKLRKQLMWLWSOTFCPLVYA---WND-----LGS 166
Db 217 RY-DRLAPGQQLLERLVGRYLDL-----APEAYARECWSDAVRRLGGS 259

RESULT 6
O12970 PRELIMINARY; PRT; 372 AA.
ID O12970
AC O12970;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Radical fringe.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIME;
RX MEDLINE=97238679; PubMed=9121551;
RA Rodriguez-Esteban C.; Schwabe J.W.; De La Pena J.; Foys B.;
RA Eshelman B.; Belmonte J.C.;
RT "Radical fringe positions the apical ectodermal ridge at the
dorsoventral boundary of the vertebrate limb.";
RL Nature 386:360-366(1997).
DR EMBL; U82088; AAC60107.1; -.
DR InterPro: IPR003378; Fringe.
DR Pfam; PF02434; Fringe; 1.
SQ SEQUENCE 372 AA; 40962 MW; 84CB7B74A4B81C6D CRC64;

Query Match 7.4%; Score 93.5; DB 13; Length 372;
Best Local Similarity 25.4%; Pred. No. 1.1;
Matches 52; Conservative 17; Mismatches 81; Indels 55; Gaps 11;

QY 4 CPSLGVTLVYLVGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDIPD--PKEKDL 61
Db 12 CILLSVTAAAVLLLLPRGQPPAAPER---RPPAPG-----SRPSPKREARPAAGSDV 61
QY 62 NETLLRSLGGHYDPGFMATSPPEDRPGGGGAAGGAEDLAEL---DQLLRQPPGAMPS 118
Db 62 -----PGDR-GGGGAAGGGRVAGSPWPSRRVRMGPPGGSAK 98
QY 119 EIKGLEFSEGLAQKKORLSKKLRKQLMWLW-----SOTFCPLVYAMNDLGSRFWRPV 173
Db 99 E--SLEKIDIFAVKTRKYKHTRLELLFQTWISRARGQTF--IFTDWEDELR-----L 149
QY 174 KVGSCFSKRSKSCS---VPEGMVCKPS 195
Db 150 KAGDHMINTCNSAVHTRQALCKMS 174

RESULT 7
O12972 PRELIMINARY; PRT; 372 AA.
ID O12972
AC O12972;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Radical fringe.
OS Gallus gallus (Chicken).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Artiosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OK NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97238680; PubMed=9121552;
 RA Lauter E., Dahn R., Orozco O.E., Yeo C.Y., Piseni J., Henrique D.,
 RA Abbot U.K., Fallon J.F., Tabin C.;
 RT "Expression of Radical Fringe in limb-bud ectoderm regulates apical
 RT "ectodermal ridge formation.";
 RL Nature 386:366-373(1997).
 DR EMBL: U91850; AAC60100.1; -
 DR InterPro: IPR003378; Fringe.
 DR Pfam: PF02434; Fringe; 1.
 SQ SEQUENCE 372 AA; 40904 MW; 05BA18254EA079D6 CRC64;
 Query Match 7.4%; Score 93.5; DB 13; Length 372;
 Best Local Similarity 25.4%; Pred. No. 1.1;
 Matches 52; Conservative 17; Mismatches 81; Indels 55; Gaps 11;
 QY 4 CSLSGTVLALVVLGRATPAGQHYLIRPAPSDNLPLVDLIEHDPIDF--PKKDL 61
 DB 12 CFLSTVTAVALLLPLPRGQPAAPRR--RPPAPG-----SRSPKREAPAGSDV 61
 QY 62 NETLRLSLGGHYDPGFMTSPEDRPGGGGAAGADLAEL--DQLRKQPSGAMS 118
 DB 62 -----PQDR-GGGSGAGGGRVAGSPWRPRRMGPPGSSAK 98
 QY 119 EIKGLEFSEGLAQGKKQRUSKRLRKLQMWLW-----SQTFCPVLYAMNDIGSRF 173
 DB 99 E--SEELKDFIAVKTTRKYHKRLLELFTQWISRAQGTF--IFTMEDRELRL----L 149
 QY 174 KVSGSCFSKRSKCS---VPEGWVCKPS 195
 DB 150 KAGDHMINTCASAVHTRQALCCWMS 174
 RESULT 8
 Q9WJ24 PRELIMINARY; PRT; 538 AA.
 ID Q9WJ24
 AC Q9WJ24
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Pr65.
 GN GAG.
 OS Moloney murine leukemia virus.
 OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
 OK NCBI_TaxID=11801;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Petropoulos C.J.;
 RT "Appendix 2 - Retroviral taxonomy, protein structure, sequences, and
 RT genetic maps.";
 RL (in) Coffin J.M. (eds.);
 RL Retroviruses, pp.757-757, Cold Spring Harbor Laboratory Press,
 RL New York (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Chappey C.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF031811; AAC82566.1; -
 DR InterPro: IPR000840; Gag_MA.
 DR InterPro: IPR002079; Gag_p12.
 DR InterPro: IPR003036; Gag_p30.
 DR InterPro: IPR002965; P_tich_extensn.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF01140; Gag_MA; 2.
 DR Pfam: PF01141; Gag_p12; 2.
 DR Pfam: PF02093; Gag_p30; 2.
 DR Pfam: PF00098; zf-CCHC; 2.
 DR PRINTS: PRO1217; PRICHEXTENSN.

DR SMART: SM00343; Znf_C2HC; 2.
 KM Core protein; Polypeptide.
 FT CHAIN 2 131 P15 MA.
 FT CHAIN 132 215 P12.
 FT CHAIN 216 478 P30 CA.
 FT CHAIN 479 534 P10 NC.
 FT CHAIN 535 538 P14 PR.
 SQ SEQUENCE 538 AA; 60858 MW; 8A7652439B464495 CRC64;
 Query Match 7.3%; Score 92; DB 15; Length 538;
 Best Local Similarity 26.5%; Pred. No. 2.3;
 Matches 45; Conservative 12; Mismatches 45; Indels 68; Gaps 9;
 QY 21 RATPAGQHY-----LHTRAP---SDN-LPLVDLI-EHPPIEPKCKDNLRTLSL 69
 DB 122 RSTPPRSSSLPALTPSLGAPKQVLSGGLDILDLTEDPYRDP----- 169
 QY 70 LGGHYDPGFMTSPEDRPGGGGA--AGADLAELDQLROR-----PSGAMS 118
 DB 170 -----PPSDRONGGEATPAGAPDPSPMASRLAGRRPPVADSTTQAPFL 217
 QY 119 EIKGLEFSEGLAQGKKQRUSKRLRKLQMWLSQTFCPVLYAMNDIGSRF 168
 DB 218 RAGG-----NGQLQYWPSSS---DLVWKKNNPSPF 245
 RESULT 9
 ID Q92808 PRELIMINARY; PRT; 1737 AA.
 AC Q92808
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Pr180.
 GN GAG-POL.
 OS Moloney murine leukemia virus.
 OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
 OK NCBI_TaxID=11801;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Petropoulos C.J.;
 RT "Appendix 2 - Retroviral taxonomy, protein structure, sequences, and
 RT genetic maps.";
 RL (in) Coffin J.M. (eds.);
 RL Retroviruses, pp.757-757, Cold Spring Harbor Laboratory Press,
 RL New York (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Chappey C.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
 CC KNOWN AS THE RETROPEPSIN FAMILY.
 DR EMBL: AF033811; AAC82568.1; -
 DR HSP: P03355; IMML.
 DR InterPro: IPR001995; Aspprotease_rtrv.
 DR InterPro: IPR001969; Aspprotease_site.
 DR InterPro: IPR000840; Gag_MA.
 DR InterPro: IPR002079; Gag_p12.
 DR InterPro: IPR003036; Gag_p30.
 DR InterPro: IPR002156; RNaseH.
 DR InterPro: IPR001584; Rye.
 DR InterPro: IPR00477; RYRbe.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF01140; Gag_MA; 2.
 DR Pfam: PF01141; Gag_p12; 2.
 DR Pfam: PF02093; Gag_p30; 2.
 DR Pfam: PR00075; rnaseh; 2.
 DR Pfam: PF00665; rye; 2.
 DR Pfam: PF00077; ryp; 2.
 DR Pfam: PF00078; ryl; 1.
 DR Pfam: PF00098; zf-CCHC; 2.
 DR SMART: SM00343; Znf_C2HC; 2.
 DR PROSITE: PS00141; ASp_PROTEASE; 2.

```
DR PROSITE; P50175; ASP_PROT_RETROV; 2.
KW Aspartyl protease; Core protein; Hydrolase; Polyprotein;
RNA-directed DNA polymerase.
FT CHAIN 2 131 P15 MA.
FT CHAIN 132 215 P12.
FT CHAIN 216 478 P10 CA.
FT CHAIN 479 534 P10 NC.
FT CHAIN 535 659 P14 PR.
FT CHAIN 660 1330 P80 RT.
FT CHAIN 1331 1737 P46 IN.
SQ SEQUENCE 1737 AA; 194782 MW; 27BB56918904E1B4 CRC64;

Query Match
Best Local Similarity 7.3%; Score 92; DB 15; Length 1737;
Matches 45; Conservative 12; Mismatches 45; Indels 68; Gaps 9;

QY 21 RATPAGQHY-----LHIRPAP---SDN-LPLVDLI-EHPDPIFPKEXDLNETLLRSL 69
DB 122 RSTPPRSSLYPALTPSLGAKPKFQVLSDSGGPLDILLTEDPFPPYRDP-----169
QY 70 LGGHYDPGFMTSPEDRPGGGGA--AGGAEDLAELDQLLR-----PSGAMPS 118
DB 170 -----PPSDRDNGGEATPAGEAPDPSPMASLRCGRREPPVADSTTSQAFPL 217
QY 119 EIKGLEFSEGLAQKKQKRLKRLQMLWSQTFPCVLYAWNGLGSRF 168
DB 218 RAGG-----NGQLQYWPFS--DLYNKNNNPSF 245

RESULT 10
ID Q42322 PRELIMINARY; PRT; 142 AA.
AC Q42322;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative IEF (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RA Tissue=GREEN SILIQUES OF A. THALIANA ECOTYPE COLUMBIA;
RA Raynal M., Greillet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z37286; CAA85542.1; -.
FT NON_TER 142 142
SQ SEQUENCE 142 AA; 15693 MW; 5BC581AF4B07A60A CRC64;

Query Match
Best Local Similarity 7.1%; Score 90; DB 10; Length 142;
Matches 27; Conservative 9; Mismatches 34; Indels 16; Gaps 4;

QY 80 ATSPEDRPGGGGAG---GAEDLAELDQLLRPSGAMPSEIKGLEFSEGLAQKKQ 135
DB 8 AVSPQATTSGGTGASPKRGKRPKTKEDS---QTPSSQQSDVRKKE-----SKKT 58
QY 136 RLKSLKRLKRLQMLWSQTFPCVLYAW 161
DB 59 QQSPSVDEKYSQW---KGLVPILYDW 81

RESULT 11
ID Q9QAP6 PRELIMINARY; PRT; 931 AA.
AC Q9QAP6;
DT 01-WAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Glycoprotein B.
```

```
OS Rangiferine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae.
OX NCBI_TaxID=79892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SALLA 82;
RX MEDLINE=99221732; PubMed=10203465;
RA Ros C., Belak S.;
RT "Studies of genetic relationships between bovine, caprine, cervine,
and rangiferine alphaherpesviruses and improved molecular methods for
virus detection and identification.";
RT J. Clin. Microbiol. 37:1247-1253(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SALLA 82;
RA Ros C., Belak S.;
RT "Characterization of the glycoprotein B gene from ruminant
alphaherpesviruses.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF078727; AAD46113.2; -.
SQ SEQUENCE 931 AA; 101624 MW; B164C67BCBC4000C CRC64;

Query Match
Best Local Similarity 7.0%; Score 88.5; DB 12; Length 931;
Matches 49; Conservative 27; Mismatches 85; Indels 59; Gaps 10;

QY 11 LVYLVVLGLRATPAGGQHYLHIRPAPSDNLPVLVDIEHPDPIFPKEXDLN-----62
DB 56 LWPLLELLPPLAPAAAAPVTAPPAPAPAP-----TEPPP-GTSPEDLDAAGNSD 109
QY 63 -ETLLRSLGGLGHYDPGFMTSPEDRPGGGGAAGGAEDLAELDQLLRPSGAMPSEIK 121
DB 110 LRAALRSQAADSESRFFVCPPP-----SCAT-----VVRLAPARPCPDYEL 151
QY 122 GLFEPSEGLAQKKQKRLKRLQMLWSQTFPCVLYAWNGLGSRF--PRYVKVGSCE 179
DB 152 GRNYTEGIGVIYKENIA-----PYTFKAVIYKKNVITTTWAGSTYAAITNQY 199
QY 180 SKRSCSVPGMGVCKPSKSVHLTVLRWC-----QRRGGOR 214
DB 200 TDR---VPVGM---AEITDVVDKKWCLSKAELRSGRR 232

RESULT 12
Q8TD19 PRELIMINARY; PRT; 979 AA.
ID Q8TD19
AC Q8TD19;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE NIMA-related kinase Nek8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Holland P.M., Milne A., Garka K., Johnson R.S., Willis C.R.,
RA Sims J.E., Rauch C.T., Bird T.A., Virca G.D.;
RT "Purification, cloning and characterization of a novel NIMA-related
kinase, Nek8, and its candidate substrate Bicd2.";
RL J. Biol. Chem. 0:0-0(2002).
DR EMBL; AY048580; AAL05428.1; -.
KW Kinase.
SQ SEQUENCE 979 AA; 107149 MW; FF2486CC599322CE CRC64;

Query Match
Best Local Similarity 7.0%; Score 88.5; DB 4; Length 979;
Matches 50; Conservative 21; Mismatches 84; Indels 67; Gaps 7;

QY 3 RCPSLGVTLYALVVLGLRATPAGGQHYLHI-----33
```

Db 625 KCGQLGVNKKRLGILNGLGPIGKQVIRVSCGDEFTIATDDNHIIFAWNGNGRLAM 684
 QY 34 ----RPAISDNLEPLVDLIEHPDPIFDP-----KENDLNETHLRLSLGG 72
 Db 685 TPTERPHGSD-----ICTSWPRPIFGSLHHVPLDLSRGWHTLILVEKVLNLSKTRNSSG 739
 QY 73 -HYDPCFMATSPEDRPPGGGGAAGAEDLAELDQLRQPSGAMPEIKGLEFSEGL-- 129
 Db 740 LSIQVYFGSSS-----PGGGGGGGGGGEEEDSQESETPD--PSGGFRGTMEADRMGELIS 793
 QY 130 ---AAGCKQRLSKLRKRLQMWLMSQTFPCVLYAMNDLGRF 168
 Db 794 PTEAMGNSNGASSSCPGWLKLENAEFIMPDPSPPLSAF 835

RESULT 13
 O8TCY4 PRELIMINARY; PRT; 979 AA.

AC O8TCY4;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE NIMA-family kinase NERCC.
 GN NERCC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Roid J., Mikhailov A., Belham C., Avruch J.;
 RT "Nercc, a mammalian NIMA-family kinase, binds the Ran GTPase and
 regulates mitotic progression.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY080896; AAL87410.1; -.
 KW Kinase.
 SQ SEQUENCE 979 AA; 107034 MW; 002483C1711BADA CRC64;

Query Match 7.0%; Score 88.5; DB 4; Length 979;
 Best Local Similarity 22.5%; Pred. No. 10;
 Matches 50; Conservative 21; Mismatches 84; Indels 67; Gaps 7;

QY 3 RCPISGVTLYALVVLGLRATRPAGGCHYH----- 33
 Db 625 KCGQLGVNKKRLGILNGLGPIGKQVIRVSCGDEFTIATDDNHIIFAWNGNGRLAM 684
 QY 34 ----RPAISDNLEPLVDLIEHPDPIFDP-----KENDLNETHLRLSLGG 72
 Db 685 TPTERPHGSD-----ICTSWPRPIFGSLHHVPLDLSRGWHTLILVEKVLNLSKTRNSSG 739
 QY 73 -HYDPCFMATSPEDRPPGGGGAAGAEDLAELDQLRQPSGAMPEIKGLEFSEGL-- 129
 Db 740 LSIQVYFGSSS-----PGGGGGGGGGGEEEDSQESETPD--PSGGFRGTMEADRMGELIS 793
 QY 130 ---AAGCKQRLSKLRKRLQMWLMSQTFPCVLYAMNDLGRF 168
 Db 794 PTEAMGNSNGASSSCPGWLKLENAEFIMPDPSPPLSAF 835

RESULT 14
 O8UCW7 PRELIMINARY; PRT; 278 AA.

AC O8UCW7;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hydrothermal protein Atu0918.
 GN Atu0918 OR AGR_C1672.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=176299;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Suthai J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.B., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Bisen J.A., Karp P.D., Boyce D. St.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Seshpimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58.";
 RL Science 294:2317-2323 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Guorillo B., Goldman B.S., Cao Y., Askenazi M., Hailing C., Mullin L.,
 RA Houniel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328 (2001).
 DR EMBL; AE009057; AAL41932.1; -.
 DR EMBL; AE008022; AAK86722.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 278 AA; 30704 MW; 73100D0C5F633D4 CRC64;

Query Match 6.9%; Score 88; DB 16; Length 278;
 Best Local Similarity 27.9%; Pred. No. 2.5;
 Matches 43; Conservative 19; Mismatches 74; Indels 18; Gaps 6;

QY 16 VVLGLRATRPAGGCHYHIRPAPSDNLEPLVDLIEHPDPIFDP-----KENDLNETHLRLSLGGHY 74
 Db 104 VVEIDETKQPARNTAPLTS---LSRLKERDSSAFPPEDALAAGRLAADPHRGL 159
 QY 75 DPGFMATSPEDRPPGGGGAAGAEDLAELDQLR---GRPSGAMPEIKGLE-----FS 126
 Db 160 NPTVATWEPDR-IASRGKEGAGALDTEAAMAAARAFRAADAMGPELSGVAIDICPE 218
 QY 127 EGIAGCKQRL---SKLRKRLQMWLMSQTFPCP 156
 Db 219 KGLETERERLWPARSAKLLRLRALSLARHYAP 252

RESULT 15
 O59129 PRELIMINARY; PRT; 814 AA.

AC O59129;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
 DE Nicotinic dehydrogenase, large chain (EC 1.5.99.4) (NDHC).
 GN NDHC.
 OS Arthrobacter nicotinovorus.
 OG Plasmid PA01.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
 OX NCBI_TaxID=29320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95115562; PubMed=7815950;
 RA Grether-Beck S., Igloi G., Pust S., Schilz E., Decker K., Brandach R.;
 RT "Structural analysis and molybdenum-dependent expression of the PA01-
 RT encoded nicotinic dehydrogenase genes of Arthrobacter nicotinovorus.";
 RL Mol. Microbiol. 13:929-936 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brandach R.;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 09:19:32 ; Search time 11 Seconds
(without alignments)
874.773 Million cell updates/sec

Title: US-09-897-322-2
Perfect score: 1268
Sequence: 1 MERCPSLGVTLYALVVVLGL.....QRCGWIPQIPIIIECKSC 232

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1268	100.0	232	1	NOGG_HUMAN	Q13253 homo sapien
2	1258	99.2	232	1	NOGG_MOUSE	P97466 mus musculus
3	1010	79.7	223	1	NOGG_CHICK	O93525 gallus gall
4	949.5	74.9	222	1	NOGG_XENLA	P49011 xenopus lae
5	782	61.7	144	1	NOGG_RAT	Q62809 rattus norv
6	759.5	59.9	223	1	NOG3_BRARE	O9yhw3 brachydanio
7	700.5	55.2	216	1	NOG1_BRARE	Q9w741 brachydanio
8	655	51.7	223	1	NOGG_FUGURU	O9yht8 fugu rubrip
9	546	43.1	212	1	NOG2_BRARE	Q9w740 brachydanio
10	96	7.6	507	1	MSI4_ARATH	O22607 arabidopsis
11	92	7.3	538	1	GAG_MLYMO	P03332 moloney mur
12	89	7.0	1093	1	SM5B_MOUSE	O60519 mus musculus
13	88	6.9	917	1	VGLB_HSVB2	P12841 bovine herp
14	87.5	6.9	226	1	TRUN_DROME	Q24155 drosophila
15	87.5	6.9	722	1	P85B_RAT	Q63788 rattus norv
16	87.5	6.9	728	1	P85B_HUMAN	O00459 homo sapien
17	86	6.8	724	1	P85B_BOVIN	P23726 bos taurus
18	85.5	6.7	315	1	PIX1_MOUSE	P70314 mus musculus
19	85.5	6.7	331	1	SHX2_MOUSE	P70390 mus musculus
20	85	6.7	867	1	SSPO_BOVIN	P98167 bos taurus
21	83	6.5	538	1	GAG_MSMYO	P03334 moloney mur
22	82.5	6.5	235	1	GAG_MSVAB	P03333 abelson mur
23	82	6.5	468	1	GAG_MSVMT	P32594 moloney mur
24	80	6.3	331	1	SHX2_HUMAN	O60902 homo sapien
25	79.5	6.3	397	1	HLK2_MALDO	O04135 malus domes
26	79.5	6.3	398	1	HLK1_MALDO	O04134 malus domes
27	79.5	6.3	1310	1	ICP4_VZVD	P09310 varicella-z
28	79	6.2	3511	1	MY15_MOUSE	O9gz24 mus musculus
29	78.5	6.2	722	1	P85B_MOUSE	O08908 mus musculus
30	78	6.2	1269	1	FLH1_HUMAN	Q13045 homo sapien
31	78	6.2	1572	1	BA12_HUMAN	O60241 homo sapien
32	77.5	6.1	388	1	SOX7_HUMAN	O9bf81 homo sapien
33	77.5	6.1	490	1	CE05_ECOLI	Q47500 escherichia

RESULT 1									
NOGG_HUMAN	1	490	1	CE10_ECOLI	Q47125 escherichia				
AC	Q13253	77.5	6.1	192	1	YK49_BPM2	O64239 mycobacteri		
DT	16-OCT-2001 (Rel. 40, Created)	76	6.0	215	1	YK47_YEAST	P28707 saccharomyc		
DT	16-OCT-2001 (Rel. 40, Last sequence update)	37	76	314	1	PIX1_HUMAN	P78337 homo sapien		
DT	16-OCT-2001 (Rel. 40, Last annotation update)	38	76	6.0	1234	1	PIP3_HUMAN	O01970 homo sapien	
DE	Noggin precursor.	39	75.5	6.0	483	1	BLAR_MELGA	P07700 meleagris g	
OS	NOG.	40	75.5	6.0	1505	1	CUT2_HUMAN	O14529 homo sapien	
GN	Homo sapiens (Human).	41	75.5	6.0	1819	1	GCP6_HUMAN	Q96rt7 homo sapien	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	42	75	5.9	391	1	SOX1_MOUSE	P53783 mus musculus	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	43	75	5.9	534	1	VGLF_MEASY	P26032 measles vir	
OX	NCBI_TaxID=9606;	44	75	5.9	559	1	PHF1_MOUSE	O921b8 mus musculus	
RN	[1]	45	74.5	5.9	287	1	BTFA_SYNY3	P72966 synechocyst	
RP	SEQUENCE FROM N.A.								
RC	TISSUE=Placenta, and Temporal cortex;								
RX	MEDLINE=9535592; PubMed=7666191;								
RA	Valenzuela D.M., Economides A.N., Rojas E., Lamb T.M., Nunez L.,								
RA	Jones P., Ip N.Y., Espinosa R. III, Brannan C.I., Gilbert D.J.,								
RA	Copeland N.G., Jenkins N.A., Le Beau M.M., Harland R.M.,								
RA	Yancopoulos G.D.;								
RT	"Identification of mammalian noggin and its expression in the adult								
RT	nervous system."								
RL	J. Neurosci. 15:6077-6084(1995).								
[2]									
VARIANTS	SYN1/SYN1 R-35; C-189; G-217; N-220; C-222; G-222 AND L-223.								
RP	MEDLINE=99178269; PubMed=10080184;								
RX	Gong Y., Krakow D., Marcelino J., Wilkin D., Chitayat D.,								
RA	Babul-Hirji R., Hudgins L., Cremers C.W., Cremers F.P.M.,								
RA	Brunner H.G., Reinker K., Rimoin D.L., Cohn D.H., Goodman F.R.,								
RA	Reardon W., Patton M., Francmano C.A., Warman M.L.;								
RT	"Heterozygous mutations in the gene encoding noggin affect human joint								
RT	morphogenesis."								
RL	Nat. Genet. 21:302-304(1999).								
CC	-!- FUNCTION: ESSENTIAL FOR CARTILAGE MORPHOGENESIS AND JOINT								
CC	FORMATION. INHIBITOR OF BONE MORPHOGENETIC PROTEINS (BMP)								
CC	SIGNALING WHICH IS REQUIRED FOR GROWTH AND PATTERNING OF THE								
CC	NEURAL TUBE AND SOMITE (BY SIMILARITY).								
CC	-!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (BY SIMILARITY).								
CC	-!- SUBCELLULAR LOCATION: Secreted.								
CC	-!- DISEASE: DEFECTS IN NOG ARE THE CAUSE OF SYMPHALANGISM PROXIMAL								
CC	SYNDROME (SYN1). SYN1 IS CHARACTERIZED BY THE HEREDITARY ABSENCE								
CC	OF THE PROXIMAL INTERPHALANGEAL (PIP) JOINTS (CUSHING								
CC	SYMPHALANGISM). SEVERITY OF PIP JOINT INVOLVEMENT DIMINISHES								
CC	TOWARDS THE RADIAL SIDE. DISTAL INTRAPHALANGEAL JOINTS ARE LESS								
CC	FREQUENTLY INVOLVED AND METACARPOPHALANGEAL JOINTS ARE RARELY								
CC	AFFECTED WHEREAS CARPAL BONE MALFORMATION AND FUSION ARE COMMON.								
CC	IN THE LOWER EXTREMITIES, TARSAL BONE COALITION IS COMMON.								
CC	CONDUCTIVE HEARING LOSS IS SEEN AND IS DUE TO FUSION OF THE STAPES								
CC	TO THE PETROSUS PART OF THE TEMPORAL BONE.								
CC	-!- DISEASE: DEFECTS IN NOG ARE THE CAUSE OF MULTIPLE SYNOSTOSES								
CC	SYNDROME 1 (SYN1). ALSO KNOWN AS SYNOSTOSES, MULTIPLE, WITH								
CC	BRACHYDACTYLIS/SYMPHALANGISM-BRACHYDACTYLIC SYNDROME. SYN1 IS								
CC	CHARACTERIZED BY TUBULAR-SHAPED (HEMICYLINDRICAL) NOSE WITH LACK								
CC	OF ALAR FLARE, OTOSCLEROTIC DEAFNESS, AND MULTIPLE PROGRESSIVE								
CC	JOINT FUSIONS COMMENCING IN THE HAND. THE JOINT FUSIONS ARE								

CC PROGRESSIVE, COMMENCING IN THE FIFTH PROXIMAL INTERPHALANGEAL
 CC JOINT IN EARLY CHILDHOOD (OR AT BIRTH IN SOME INDIVIDUALS) AND
 CC PROGRESSING IN AN UNWEAR-TO-RADIAL AND PROXIMAL-TO-DISTAL
 CC DIRECTION, WITH INCREASING AGE, ANKYLOSIS OF OTHER JOINTS,
 CC INCLUDING THE CERVICAL VERTEBRAE, HIPS, AND HUMERORADIAL JOINTS,
 CC DEVELOP.

CC -1- SIMILARITY: BELONGS TO THE NOGIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U31202; AAA83259.1; -.

DR Genew; HGNC:7866; NOG.

DR MIM; 602991; -.

DR MIM; 185800; -.

DR MIM; 185500; -.

DR Glycoprotein; Signal; Disease mutation.

KW SIGNAL 1 27 POTENTIAL.

FT CHAIN 28 232 NOGIN.

FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARIANT 35 35 P -> R (IN SYM1).

FT VARIANT 189 189 /FTID=VAR.011361.

FT VARIANT 217 217 G -> C (IN SYM1).

FT VARIANT 217 217 /FTID=VAR.011362.

FT VARIANT 220 220 W -> G (IN SYM1).

FT VARIANT 222 222 /FTID=VAR.011363.

FT VARIANT 222 222 I -> N (IN SYM1).

FT VARIANT 222 222 /FTID=VAR.011364.

FT VARIANT 222 222 Y -> C (IN SYM1).

FT VARIANT 222 222 /FTID=VAR.011365.

FT VARIANT 222 222 Y -> D (IN SYM1).

FT VARIANT 223 223 /FTID=VAR.011366.

FT VARIANT 223 223 P -> L (IN SYM1).

FT SEQUENCE 232 AA; 25774 MW; FCAOD8E1E2098580 CRC64;

SQ

Query Match 100.0%; Score 1268; DB 1; Length 232;
 Best Local Similarity 100.0%; Pred. No. 1.8e-106;
 Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERCPSELGVTLYALVVLGLRATPAGGQHYLHIRAPSDNLPLVDLIHPPIDFPKXED 60
 DB 1 MERCPSELGVTLYALVVLGLRATPAGGQHYLHIRAPSDNLPLVDLIHPPIDFPKXED 60

QY 61 LNETLLRSLLGGHYDPGFMTSPEDRPGGGGGAAGABDLAEIDOLLRRPSGAMPSSEI 120
 DB 61 LNETLLRSLLGGHYDPGFMTSPEDRPGGGGGAAGABDLAEIDOLLRRPSGAMPSSEI 120

QY 121 KGLEFSEGLAOGKKORLSKRLQMWLMSTQFCPLVYAMNDLGSFMPRYVVGSCFS 180
 DB 121 KGLEFSEGLAOGKKORLSKRLQMWLMSTQFCPLVYAMNDLGSFMPRYVVGSCFS 180

QY 121 KGLEFSEGLAOGKKORLSKRLQMWLMSTQFCPLVYAMNDLGSFMPRYVVGSCFS 180
 DB 121 KGLEFSEGLAOGKKORLSKRLQMWLMSTQFCPLVYAMNDLGSFMPRYVVGSCFS 180

QY 181 KRCSCVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPIQYPIISECKSC 232
 DB 181 KRCSCVPEGMVCKPSKSVHLTVLRMRCQRRGRCGMIPIQYPIISECKSC 232

RESULT 2
 NOGGIN_MOUSE
 ID NOGGIN_MOUSE STANDARD; PRT; 232 AA.
 AC P97466;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Noggin precursor.
 GN NOG.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxId=10090;
 CC RN
 CC SEQUENCE FROM N.A.
 CC RC STRAIN=129/5v;
 CC RX MEDLINE=98252829; PubMed=9585504;
 CC RA McMahon J.A., Takada S., Zimmerman L.B., Fan C.-M., Harland R.M.,
 CC McMahon A.P.;
 CC RT "Noggin-mediated antagonism of BMP signaling is required for growth
 CC and patterning of the neural tube and somite.";
 CC RL Genes Dev. 12:1438-1452(1998).
 CC RN [2]
 CC RP FUNCTION.
 CC RX MEDLINE=98267313; PubMed=9603738;
 CC RA Brunet L.J., McMahon J.A., McMahon A.P., Harland R.M.;
 CC RT "Noggin, cartilage morphogenesis, and joint formation in the mammalian
 CC skeleton.";
 CC RL Science 280:1455-1457(1998).
 CC CC -1- FUNCTION: ESSENTIAL FOR CARTILAGE MORPHOGENESIS AND JOINT
 CC FORMATION. INHIBITOR OF BONE MORPHOGENETIC PROTEINS (BMP)
 CC SIGNALING WHICH IS REQUIRED FOR GROWTH AND PATTERNING OF THE
 CC NEURAL TUBE AND SOMITE.
 CC CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (BY SIMILARITY).
 CC CC -1- SUBCELLULAR LOCATION: Secreted.
 CC CC -1- TISSUE SPECIFICITY: EXPRESSED IN CONDENSING CARTILAGE AND IMMATURE
 CC CHONDROCYTES.
 CC CC -1- DEVELOPMENTAL STAGE: EMBRYONIC EXPRESSION WAS FIRST DETECTED IN
 CC THE NOBE AT 7.5 DAYS POSTCOITUM. BY EARLY SOMITE STAGES,
 CC EXPRESSION EXTENDS ANTERIORLY ALONG THE ENTIRE LENGTH OF THE
 CC NOTOCHORD AND IS EXPRESSED IN THE DORSAL NEURAL TUBE FROM THE
 CC CAUDAL HINDBRAIN TO THE POSTERIOR-MOST REGION OF THE EMBRYO. BY
 CC THE TIME CRANIAL TUBE CLOSURE IS COMPLETED EXPRESSION IS
 CC CONTINUOUS ALONG MOST OF THE DORSAL MIDLINE OF THE NEURAL TUBE. TO
 CC ITS ROSTRAL TERMINATION AT THE BASE OF THE FOREBRAIN. EXPRESSION
 CC IN THE NEURAL TUBE AND CAUDAL NOTOCHORD REMAINS UNCHANGED DURING
 CC EARLY ORGANOGENESIS FROM 9.5 DPC TO 10.5 DPC.
 CC CC -1- DISEASE: DEFECTS IN NOG ARE THE CAUSE OF A RECESSIVE LETHAL
 CC PHENOTYPE AT BIRTH. MULTIPLE DEFECTS INCLUDE A FAILURE OF NEURAL
 CC TUBE CLOSURE, BROAD CLUB-SHAPED LIMBS, LOSS OF CAUDAL VERTEBRAE, A
 CC SHORTENED BODY AXIS, AND RETENTION OF A SMALL VESTIGIAL TAIL.
 CC CC -1- SIMILARITY: BELONGS TO THE NOGIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U79163; AAB38281.1; -.

DR MGD; MGI:104327; NOG.

DR Glycoprotein; Signal.

FT SIGNAL 1 27 POTENTIAL.

FT CHAIN 28 232 NOGIN.

FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 232 AA; 25770 MW; 0192AB6373B78B74 CRC64;

SQ

Query Match 99.2%; Score 1258; DB 1; Length 232;
 Best Local Similarity 99.1%; Pred. No. 1.4e-105;
 Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MERCPSELGVTLYALVVLGLRATPAGGQHYLHIRAPSDNLPLVDLIHPPIDFPKXED 60
 DB 1 MERCPSELGVTLYALVVLGLRATPAGGQHYLHIRAPSDNLPLVDLIHPPIDFPKXED 60

QY 61 LNETLLRSLLGGHYDPGFMTSPEDRPGGGGGAAGABDLAEIDOLLRRPSGAMPSSEI 120
 DB 61 LNETLLRSLLGGHYDPGFMTSPEDRPGGGGGAAGABDLAEIDOLLRRPSGAMPSSEI 120

QY 121 KGLEFSEGLAOGKKORLSKRLQMWLMSTQFCPLVYAMNDLGSFMPRYVVGSCFS 180
 DB 121 KGLEFSEGLAOGKKORLSKRLQMWLMSTQFCPLVYAMNDLGSFMPRYVVGSCFS 180


```

CC -!- SIMILARITY: BELONGS TO THE NOGIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF095337; AAC72965.1; -.
KW Glycoprotein; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 223 NOGIN.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 223 AA; 25121 MW; 35CE122E1B27F729 CRC64;

Query Match
Best Local Similarity 53.7%; Pred. No. 1,3e-51;
Matches 123; Conservative 37; Mismatches 57; Indels 12; Gaps 5;

QY 5 PSIGVTLYALVVLGLRATPAGGQHYLRAPSDNLPVLDLIEHPDIPDKEDINET 64
DB 4 PRLRVATYLLLSVGLLHGACOPYLLLRIPEDSLPIVELKEDPQVPDPKRDINET 63
QY 65 LRLSLGLGHYDPGFMAISPPEDRRGGGGAAGADLAELDQLLRQPSGAMPEIKGLE 124
DB 64 ELKSVL-GPDFSRFLSVLPAAE-----DGHAGN---DELDDFDAQWGGALPKEIRAVD 113
QY 125 FSEGLAOGKKORLSKRLKRLQMWLMSQTFCPVLYAMNDLGSFMPRYVKGSCFSKRSK 184
DB 114 F-DAPQLGKKHKSFKLKLQMLWALYSFCPLAHAMTDLGSRRPFRVAAAGSLSGRSK 172
QY 185 SVPEGWCKPSKSVHLTVLRMRC-QRRGQRCGWIPIQYPIISCKKSC 232
DB 173 SVPEGWCKPRASTHLLTILRMRCVQRKVGKLCAMIPMQYPIITDCKSC 221

RESULT 9
NOG2_BRARE
AC 09W740;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Noggin 2 precursor.
GN NOG2.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=99423658; PubMed=10491267;
RA Fuerthner M., Thiesse B., Thiesse C.;
RT "Three different noggin genes antagonize the activity of bone
RT morphogenetic proteins in the zebrafish embryo."
RL Dev. Biol. 214:181-196(1999).
CC -!- FUNCTION: INHIBITOR OF BONE MORPHOGENETIC PROTEINS (BMP)
CC SIGNALING.
CC -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS AT THE END OF GASTRULATION IN
CC THE AXIAL MESODERM. BY THE 5-SOMITE STAGE, EXPRESSED IN
CC TELENCEPHALON AND ANTERIOR DIENCEPHALON. FROM EARLY SEGMENTATION
CC STAGES UNTIL THE END OF TAIL ELONGATION, FOUND IN THE TAIL BUD.
CC EXPRESSION IS MAINTAINED DURING SCOTIOGENESIS. AT THE 10-SOMITE
CC STAGE, DETECTED IN THE ANTERO-MEDIAL ASPECT OF THE SOMITES. AT 20
CC HRS OF DEVELOPMENT EXPRESSION IS OBSERVED IN THE VENTRO-MEDIAL
CC PART OF THE SOMITE AS WELL AS IN A SMALL POPULATION OF CELLS
CC LOCATED MORE DORSALLY, ADJACENT TO THE NEURAL TUBE. AS DEVELOPMENT
CC PROCEEDS, EXPRESSION IN THE DORSAL SOMITE IS PROGRESSIVELY LOST,

```

```

CC WHILE VENTRO-MEDIAL SCLEROTOMAL CELLS LINING THE DEVELOPING AXIAL
CC VASCULATURE CONTINUE TO SHOW EXPRESSION UNTIL 30 HRS OF
CC DEVELOPMENT. IN THE HEAD, EXPRESSION IS MAINTAINED IN THE
CC TELENCEPHALON AND ANTERIOR DIENCEPHALON UNTIL LATE EMBRYOGENESIS.
CC AT THE 15-SOMITE STAGE, EXPRESSED IN THE FOREBRAIN, DORSAL
CC HINDBRAIN AND DORSAL CAUDAL SPINAL CORD. UNTIL LATE STAGES OF
CC EMBRYOGENESIS STRONG EXPRESSION IS OBSERVED IN THE DORSAL
CC HINDBRAIN WALLS.
CC -----
CC -!- SIMILARITY: BELONGS TO THE NOGIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF159148; AAD43133.1; -.
KW Glycoprotein; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 212 NOGIN 2.
FT CARBOHYD 84 84 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 212 AA; 24531 MW; D0262DD09B9D6504 CRC64;

Query Match
Best Local Similarity 43.1%; Score 546; DB 1; Length 212;
Matches 107; Conservative 30; Mismatches 56; Indels 36; Gaps 4;

QY 14 LVVVLGLRATPAGGQHYLRAPSDNLPVLDLIEHPDIPDKEDINETLLSLGLGH 73
DB 10 LLLLGLCAHGTASQHLRLRLPSSGFLVPLINPPEHDPKRDISEKTLKKGSN 69
QY 74 YDPGFV-----ATSPEDRRPGGCGGAAGADLAELDQLLRQPSGAMPEIKGLE 124
DB 70 FDNFMSHLPQNLMSAPE-----LRLMPMAELKLD 104
QY 125 FSEGLAOGKKORLSKRLKRLQMWLMSQTFCPVLYAMNDLGSFMPRYVKGSCFSKRSK 184
DB 105 LIE-TFYGRRVKVKKARRKFLQMLWMTYHCPLVLTWKDLGLRFMPRYIKENGCFSKRSK 163
QY 185 SVPEGWCKPSKSVHLTVLRMRCQRRGQRCGWIPIQYPIISCKKSC 232
DB 164 SPPEGWCKPRAVAVYKTFRLWYCGFMROKCTWTIQVPIISCKKSC 212

RESULT 10
MS14_ARATH
AC 022607; Q9SLD1; Q93VF7; Q42322; Q42323;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE WD-40 repeat protein MS14.
GN MS14 OR AT2G19520 OR F3P11.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN 1;
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.D., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;

```

Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.";
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Riken Arabidopsis full length cDNA clones (RAPs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 4-507 FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=99087420; PubMed=9872415;
 RA Kenzior A.L., Folk W.R.;
 RT "AtMS14 and RbAp48 WD-40 repeat proteins bind metal ions.";
 RL FEBS Lett. 440:425-429(1998).
 RN [4]
 RP SEQUENCE OF 1-142 AND 452-507 FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Green silique;
 RA Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 CC -!- SIMILARITY: BELONGS TO THE RBAP46/RBAP48/MS11 FAMILY OF WD-REPEAT
 CC PROTEINS.
 CC -!- CAUTION: Ref.4 (CAA85542) sequence differs from that shown due to
 CC a frameshift in position 136.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AC005917; AAD10151.2; ALT_INIT.
 DR EMBL; AY059799; AAL24281.1; -.
 DR EMBL; AY057655; AAL15286.1; -.
 DR EMBL; AY081447; AAM10009.1; -.
 DR EMBL; AF028711; AAD03340.1; -.
 DR EMBL; Z37286; CAA85542.1; ALT_FRAME.
 DR EMBL; Z37287; CAA85543.1; -.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 6.
 DR PRINTS; PR00320; GPROTEINERPT.
 DR ProDom; PD000018; WD40; 2.
 DR SMART; SM00320; WD40; 5.
 DR PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
 DR PROSITE; PS00678; WD_REPEATS_2; 3.
 DR PROSITE; PS00678; WD_REPEATS_3; 3.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat; Nuclear protein.
 FT REPEAT 162 193 WD 1.
 FT REPEAT 217 248 WD 2.
 FT REPEAT 290 321 WD 3.
 FT REPEAT 335 366 WD 4.
 FT REPEAT 384 415 WD 5.
 FT REPEAT 439 477 WD 6.
 FT DOMAIN 468 471 POLY-GLY.
 FT CONFLICT 89 89 W -> L (IN REF. 4; CAA85542).
 FT CONFLICT 126 126 V -> F (IN REF. 4; CAA85542).
 FT CONFLICT 202 202 A -> P (IN REF. 3).
 FT CONFLICT 270 270 T -> P (IN REF. 3).
 FT CONFLICT 463 463 D -> A (IN REF. 4; CAA85543).
 FT CONFLICT 489 489 V -> F (IN REF. 4; CAA85543).
 FT CONFLICT 494 494 E -> A (IN REF. 3).
 SQ SEQUENCE 507 AA; 55759 MW; C37F800F8B33397 CRC64;
 Query Match 7.6%; Score 96; DB 1; Length 507;
 Best Local Similarity 30.5%; Pred. NO. 0.38;
 Matches 29; Conservative 10; Mismatches 40; Indels 16; Gaps 4;

QY 80 ATSPPPDRGGGGAAG----GAEDLAEIDLQRLRQPPSGAMPSEIKGLSEGLAQKKQ 135
 DB 8 AVSPQATTPSGGTGASGPKRGKPKTKEDS---QTPSSQQQSDVRWKE-----SGKKT 58
 QY 136 RLSSKLLRRKLQWLWSQTFCPLVYAWNDLGSFRWP 170
 DB 59 QQSPSVDEKYSQW---KGLVPILYDMLANHNLVWP 90
 RESULT 11
 ID _GAG_MLVMO STANDARD; PRT; 538 AA.
 AC F03332;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GAG polyprotein [Contains: Core protein P15; Inner coat protein P12;
 DE Core shell protein P30; Nucleoprotein P10].
 OS Moloney murine leukemia virus.
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11801;
 RN [1] _SEQUENCE FROM N.A. (CLONE PMLV-1).
 RP MEDLINE=82035843; PubMed=6169994;
 RA Shinnick T.M., Lerner R.A., Sutcliffe J.G.;
 RT "Nucleotide sequence of Moloney murine leukaemia virus.";
 RL Nature 293:543-548(1981).
 RN [2]
 RP SEQUENCE OF 2-31, AND MYRISTOYLATION.
 RX MEDLINE=83169654; PubMed=6340098;
 RA Henderson L.E., Krutzsch H.C., Oroszlan S.;
 RT "Myristyl amino-terminal acylation of murine retrovirus proteins: an
 RT unusual post-translational proteins modification.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:339-343(1983).
 RN [3]
 RP SEQUENCE OF 479-529.
 RX MEDLINE=81264245; PubMed=6267042;
 RA Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W.,
 RA Oroszlan S.;
 RT "Primary structure of the low molecular weight nucleic acid-binding
 RT proteins of murine leukemia viruses.";
 RL J. Biol. Chem. 256:8400-8406(1981).
 CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
 CC POLYPROTEIN.
 CC -!- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J02255; AAB59942.1; -.
 DR PIR; A03930; FOMVIM.
 DR InterPro; IPR000840; Gag_MA.
 DR InterPro; IPR002079; Gag_P12.
 DR InterPro; IPR003036; Gag_P30.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF00098; zf-CCHC; 1.
 DR Pfam; PF01140; Gag_MA; 1.
 DR Pfam; PF01141; Gag_P12; 1.
 DR Pfam; PF02093; Gag_P30; 1.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS0158; ZF_CCHC; 1.
 KW Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;
 KW Zinc-finger.
 FT CHAIN 2 131 CORE PROTEIN P15.
 FT CHAIN 132 215 INNER COAT PROTEIN P12.
 FT CHAIN 216 478 CORE SHELL PROTEIN P30.
 FT CHAIN 479 534 NUCLEOPROTEIN P10.

```

Query Match: 7.3%; Score 92; DB 1; Length 538;
Best Local Similarity 26.5%; Pred. No. 0.93;
Matches 45; Conservative 12; Mismatches 45; Indels 60; Gaps 9;

QY 21 RATPAGGQHY-----LHIRPAP--SDN-LPLVDLI-EHPDIFDPKEKDLNETLLNSL 69
   |||
Db 122 RSTPPRSSLYPALTPSLGAKPKQVLDSDGGPLIDLTTEDPPRYRDR----- 169
QY 70 LCGHIDPCGMATSPPEPDRCGGGGA--AGCAEDLALDQLLOR-----PSGAMS 118
   |||
Db 170 -----PPPSDRDONGGEATPAGEADPPSPMASRLRRRPPVADSTTSGAFPL 217
QY 119 EIKGLEFSGLAQGKKQRLSKLRRTLQMWLWQTCFPLVYAMNDGSRF 168
Db 218 RAGG-----NGQLQIWPSSS---DLYWKNNPSP 245

RESULT 12
SMSB_MOUSE STANDARD; PRT: 1093 AA.
ID SMSB_MOUSE
AC Q60519;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Semaphorin 5B precursor (Semaphorin G) (Sema G).
GN SEMA5B OR SEMAG OR SEMG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI;
RA MEDLINE=96414430; PubMed=8817451;
RA Adams R.H., Betz H., Pueschel A.W.;
RT "A novel class of murine semaphorins with homology to thrombospondin
RL Mech. Dev. 57:33-45(1996).
CC -1- FUNCTION: MAY ACT AS POSITIVE AXONAL GUIDANCE CUES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: IN ADULT, ONLY DETECTED IN BRAIN.
CC -1- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED IN EMBRYONIC AND
CC ADULT TISSUES. ITS ABUNDANCE DECREASES FROM E10 TO B1TH.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/).license@isb-sib.ch).
-----
CC EMBL, X97818, CAA66398.1; -.
DR MGD; MGI:107555; Sema5b.
DR InterPro; IPR003659; Plexin-like.
DR InterPro; IPR002165; Plexin_repeat.
DR InterPro; IPR001627; Sema.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF01403; Sema; 1.
DR Pfam; PF01437; PSI; 1.
DR SMART; SMO0423; PSI; 1.
DR SMART; SMO0209; TSP1; 4.
DR PROSITE; PS50092; TSP1, 6.
DR Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;
DR Developmental protein; Glycoprotein.

```

FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	1093	SEMAPHOTIN 5B.
FT	DOMAIN	20	978	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	979	999	POTENTIAL.
FT	DOMAIN	1000	1093	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	236	518	SEMA.
FT	DOMAIN	551	605	TSP TYPE-1 1.
FT	DOMAIN	606	663	TSP TYPE-1 2.
FT	DOMAIN	664	720	TSP TYPE-1 3.
FT	DOMAIN	721	776	TSP TYPE-1 4.
FT	DOMAIN	795	851	TSP TYPE-1 5.
FT	DOMAIN	852	908	TSP TYPE-1 6.
FT	DOMAIN	909	958	TSP TYPE-1 7.
FT	CARBOHYD	59	59	N-LINKED (GLCNAC. .)
FT	CARBOHYD	95	95	(POTENTIAL).
FT	CARBOHYD	157	157	N-LINKED (GLCNAC. .)
FT	CARBOHYD	178	178	(POTENTIAL).
FT	CARBOHYD	287	287	N-LINKED (GLCNAC. .)
FT	CARBOHYD	333	333	(POTENTIAL).
FT	CARBOHYD	378	378	N-LINKED (GLCNAC. .)
FT	CARBOHYD	532	532	(POTENTIAL).
FT	CARBOHYD	539	539	N-LINKED (GLCNAC. .)
FT	CARBOHYD	547	547	(POTENTIAL).
FT	CARBOHYD	602	602	N-LINKED (GLCNAC. .)
FT	CARBOHYD	728	728	N-LINKED (GLCNAC. .)
FT	CARBOHYD	944	944	(POTENTIAL).
SO	SEQUENCE	1093 AA;	120326 NM;	29ECC9B1E8108717 CRC64;

Query Match 7.0% Score 89; DB 1; Length 1093;
Best Local Similarity 23.8%; Pred. No. 4;
Matches 57; Conservative 22; Mismatches 78; Indels 82; Gaps 15.

QY	43	IVDLIHPDPFPDPKX-----KDINETLRLSLGSHYDPGMATSPPEDRPGGGG	93
Dd	417	VVDIVAGADTLVHVLVIGTESGLIKALS-TSRSLRCYLEE--LVHLF-----	464
QY	94	AAGGADELAEI-----DOLLRORPCAMPSETIKGLEFSGLAOK-----	133
Dd	465	--GRLEPFRSLRIHSARALFYGLSDRYLR-----IPLECSAYHSOGACLGARDPYCG	516
QY	134	---KRLSKULRKIQOMLSQ--FFCVLVAMNDLSGRFPRVY-----KYSGCPS	180
Dd	517	WDGKKQLCSLTLEDSSNMSLMITTCVLRNVYTRDGGFGFWSPMKPCSHLDGDNSGSLC	576
QY	181	K-RSCSVE---GWVCPSKSVHLTVLRMRCQRGRC--GWIPIQVPIISECKSC	232
Dd	577	RARSCDSPRRPGGLECL-GPSIHIA----NCNRKGATMTASSW-----AQCTSTC	622

RESULT 13
VGBL_HSVB2
ID_VGBL_HSVB2 STANDARD; PRT; 917 AA.
AC PI2641;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein B-1 precursor.
OS Bovine herpesvirus type 2 (strain BMV) (Bovine mammillitis virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OX Alphaherpesvirinae; Simplexvirus.
XC NCBI_TaxID=10296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306231; PubMed=2841793;
RA Hammerschmidt W., Contrats F., Menkertz J., Pauli G., Ludwig H.,
RA Buik H.-J.;
RT "Conservation of a gene cluster including glycoprotein B in bovine
RT herpesvirus type 2 (BNV-2) and herpes simplex virus type 1 (HSV-1).";
RL Virology 165:388-405 (1988).
[2]
RP SEQUENCE OF 1-200 FROM N.A.
RX MEDLINE=88306232; PubMed=2457278;
RA Hammerschmidt W., Contrats F., Menkertz J., Buik H.-J., Pauli G.,

```

RA Ludwig H.;
RT "Common epitopes of glycoprotein B map within the major DNA-binding
RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT virus type 1 (HSV-1).";
RL Virology 165:406-418(1988).
CC -!- FUNCTION: GB1 IS A 130 kDa GLYCOPROTEIN WHICH IS NECESSARY FOR THE
CC PENETRATION OF THE VIRUS INTO THE HOST CELL AND THE INDUCTION OF A
CC SYNCYTIAL PHENOTYPE.
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M21628; AAA46053.1; -.
DR EMBL; M21632; AAA46052.1; -.
DR PIR; C29242; VGBEBH.
DR InterPro; IPR000234; Glycoprot B.
DR Pfam; PF00606; Glycoprotein B; 1.
DR ProDom; PD000693; Glycoprot_B; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 917 GLYCOPROTEIN B-1.
FT TRANSMEM 578 594 POTENTIAL.
FT TRANSMEM 770 786 POTENTIAL.
FT TRANSMEM 795 811 POTENTIAL.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 164 164 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 692 692 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 917 AA; 101882 MW; 1B96CBF50DB4D3F3 CRC64;

Query Match 6.9%; Score 88; DB 1; Length 917;
Best Local Similarity 23.9%; Pred. No. 4;
Matches 42; Conservative 16; Mismatches 52; Indels 66; Gaps 6;

QY 38 SDNLPVLDLIEHPDPFDPKEDLNELT-----PMEKEEDPTLAPRASRDAPGTPKVPAMPGVTPPSGNASEPA 115
Db 61 SPSLPALDITPQP-----PMEKEEDPTLAPRASRDAPGTPKVPAMPGVTPPSGNASEPA 115
QY 66 -----LRSLGLGHYDPCFMATSPEDRPGGGGAAGAEADLAELDQLLRQPSGAMP 117
Db 116 DPALRADLRGLKSSDDPNFYCPPTG-----ATVRLLEPRPCPELP 160
QY 118 SEIKGLEFSEGLAQGKKQRSLKRLKQLWLSQTFCPVL--YAWNDLGSRWPR 171
Db 161 ---KGLNFTGEGIAVTEKENLA-PYKFKATYYKAVTVASVWSGYSYNQFMNIFEDR 212

RESULT 14
TRUN DROME
ID TRUN DROME STANDARD; PRT; 226 AA.
AC Q24155; Q9VL09;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Trunk protein precursor.
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN TRK OR CG5619.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=96033803; PubMed=7590233;
RA Casanova J., Furtiols M., McCormick C.A., Struhl G.;
RT "Similarities between trunk and spatze, putative extracellular
RT ligands specifying body pattern in Drosophila.";
RL Genes Dev. 9:2539-2544(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: REQUIRED FOR ACTIVITY OF THE TOR RECEPTOR, COULD BE
CC A LIGAND OF TOR. INVOLVED IN SPECIFYING TERMINAL BODY PATTERN.
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41064; AAC46953.1; -.
DR EMBL; AB003628; AAF52896.1; ALT_INIT.
DR FlyBase; FBgn003751; trk.
KW Developmental protein; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 226 TRUNK PROTEIN.
SQ SEQUENCE 226 AA; 26348 MW; 2CF1B5049C63521B CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 226;
Best Local Similarity 25.4%; Pred. No. 0.84;
Matches 44; Conservative 22; Mismatches 58; Indels 49; Gaps 10;

QY 60 DLNETLLRSILGLGHYDPGFMTSPEDRPGGGGAAGAEADLAELDQLLRQPSGAMP-- 117

```



```

Db      29 ELSTOGLAKLGFAPNRVMSIDP-----GEPEKSY--HLGYRSSYELPFY 75
Qy      118 -----SEIKLEPSE-GIAQCKOR-LSKILR-----KIQMLMS 151.
Db      76 ADSDAISVSHFPAWETHPALVEKKEAPRSKSLRTSAPFMDRYGHPRIDGFKORPW--- 132
Qy      152 QTFCEPLVAMNDGSRFWPVRVYKVGSCFSCSRSCVPEGMWCKPSKVLTLR 204
Db      133 --ECSSKINMIDGLNFPYRISIECIA-RKMW-DHFNCKP-KSFTIKVL 180

RESULT 15
P85B_RAT
ID_P85B_RAT STANDARD; PRT; 722 AA.
AC Q63788;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphatidylinositol 3-kinase regulatory beta subunit (PI3-kinase
p85-beta subunit) (Pcdins-3-kinase p85-beta).
GN PIK3R2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RX MEDLINE=96214979; PubMed=8621382;
RA Inukai K., Anai M., Vandreva E., Hosaka T., Katagiri H., Funaki M.,
RA Fukushima Y., Ogihara T., Yazaki Y., Kikuchi M., Oka Y., Amano T.,
RT "A novel 55-kDa regulatory subunit for phosphatidylinositol 3-kinase
RT of the p85alpha gene."
RL J. Biol. Chem. 271:5317-5320(1996).
CC -!- FUNCTION: BINDS TO ACTIVATED (PHOSPHORYLATED) PROTEIN-TYROSINE
CC KINASES, THROUGH ITS SH2 DOMAIN, AND ACTS AS AN ADAPTER, MEDIATING
CC THE ASSOCIATION OF THE P110 CATALYTIC UNIT TO THE PLASMA MEMBRANE.
CC -!- SUBUNIT: HETERODIMER OF A P110 (CATALYTIC) AND A P85 (REGULATORY)
CC SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE PI3K P85 SUBUNIT FAMILY.
CC -!- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
Db      EMBL: D64046; BAA10926.1; -
Db      HSSP: P23727; 2PNB.
Db      InterPro: IPR001720; PI3kinase_P85.
Db      InterPro: IPR000198; RhOGAP.
Db      InterPro: IPR000980; SH2.
Db      InterPro: IPR001452; SH3.
Db      Pfam: PF00017; SH2; 2.
Db      Pfam: PF00018; SH3; 1.
Db      Pfam: PF00620; RhOGAP; 1.
Db      PRINTS: PR00678; PI3KINSEP85.
Db      PRINTS: PR00401; SH2DOMAIN.
Db      ProDom: PD000066; SH3; 1.
Db      ProDom: PD000093; SH2; 2.
Db      SMART: SM00324; RhOGAP; 1.
Db      SMART: SM00325; SH2; 2.
Db      SMART: SM00326; SH3; 1.
Db      PROSITE: PS50001; SH2; 2.
Db      PROSITE: PS50002; SH3; 1.
Db      SH3 domain; SH2 domain; Repeat.

```

```

FT      DOMAIN 4 80 SH3.
FT      DOMAIN 125 257 RHO-GAP.
FT      DOMAIN 324 419 SH2 1.
FT      DOMAIN 616 710 SH2 2.
SQ      SEQUENCE 722 AA; 81328 MW; 1208368B9F6FC95 CRC64;

Query Match 6.9%; Score 87.5; DB 1; Length 722;
Best Local Similarity 25.0%; Pred. No. 3.3;
Matches 46; Conservative 22; Mismatches 63; Indels 53; Gaps 10;

Qy      3 RCPSLGVTALVALVVLG---LRATPAGQHYLHIRPAPSDNPLVLDLIEHPDIPDKER 59
Db      235 RAPSPATAVAHALASAFGPLLIRAPPGEGD--GSEBPAP----- 271
Qy      60 DINEETLRLSLGGHYD-----PGFMATSPEDRPG---GGGAAGAGADL----- 101
Db      272 DFPVLLIERLVQEHVDQDTAPPALPKPSKVPAPFALTALANGSPPSLQDAEWYGDISR 331
Qy      102 AEIDQLRQRPAGM-----PSRIKGLFSEGGIAQCKORLSKILRKQLQMLMSQ--TF 154
Db      332 EEWNERLRDTPDGTFLVRDASSKIQG-EYTLIRKGNKKLIKVFHRD-GHYGFSPLTF 389
Qy      155 CPVL 158
Db      390 CSVV 393

```

Search completed: January 7, 2003, 09:23:53.
Job time : 14 secs


```

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84577
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-507 <STO>
A:Cross-references: GB:AE002093; NID:g4191782; PIDN:AAD10151.1; GSPDB:GN00139
C:Genetics:
A:Gene: Atg919520
A:Map position: 2

Query Match 7.6%; Score 96; DB 2; Length 507;
Best Local Similarity 30.5%; Pred. No. 0.69;
Matches 29; Conservative 10; Mismatches 40; Indels 16; Gaps 4;

Qy 80 ATSPEDRPGCGGAG---GAEDLAELDQLRQRPSPGMPSEIKLEFSEGLAOGKKQ 135
Db 8 AVSPQATTPSGGTGASGPKRGRKPKTKEDS---QTPSSQQGSDVKMKE-----SGKKT 58
Qy 136 RLKKLRKRLQMLMSQTFPCVLVYAMNDGSRWP 170
Db 59 QQSPSVDEKYSQW---KGLVPLTYDMLANHNLVMP 90

RESULT 3
139595
phospholipase C - Aeromonas hydrophila
C:Species: Aeromonas hydrophila
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 08-Oct-1999
C:Accession: I39595
R:Ingham, A.B.; Pemberton, J.M.
Curr. Microbiol. 31, 28-33, 1995
A>Title: A lipase of Aeromonas hydrophila showing nonhemolytic phospholipase C activity
A:Reference number: I39595; MUID:95284718; PMID:7767226
A:Accession: I39595
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-684 <RES>
A:Cross-references: EMBL:U14011; NID:9537628; PIDN:AAA75598.1; PID:9537629

Query Match 7.5%; Score 94.5; DB 2; Length 684;
Best Local Similarity 25.2%; Pred. No. 1.3;
Matches 58; Conservative 22; Mismatches 89; Indels 61; Gaps 10;

Qy 19 GLRATPAGCGOYHLIRPAPSDNLPLVDLIEHPDIFDPKEDINETLLRSILGG--HYD 75
Db 289 GSRAVPASGAKDA-IATAVAPSRPLAELREPRFIPGRSSCLTSSPRLAELVPGIKRPS 347
Qy 76 PGPMTATP-----PEDRPG---GGG-----AAGAEPLAELDQ 106
Db 348 PGTVPSPVRRHROCAEGVTSDRRPGRGCGSLAWGTDGRSDSPRAAGGEGAHRGDP 407
Qy 107 LLRORPGAMSEIKLEFSEGLAOGKKORLSKKLRLKQLMWLMSQTFCVLYAMNDL-- 164
Db 408 HLRGQAGSR-AETRPDLPPAAGAGGATPADAHLCRRAV---CHHRCHLSSMKRLGE 462
Qy 165 GSRFWPYYVVKVSGSCFSKRSQSVPEGVCKPSKSVHLTVLRMQRQRRGOR 214
Db 463 GERLCP-----GAGPDRCRCSKHHGQEC-----RRGDR 491

RESULT 4
FOMVIM
gag polyprotein - Moloney murine leukemia virus
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein
C:Species: Moloney murine leukemia virus
A>Note: host Mus sp. (mouse)
C:Date: 01-Sep-1981 #sequence_revision 27-Nov-1985 #text_change 24-Jul-1997
C:Accession: A03930
R:Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.
Nucleic Acids Res. 9, 119-126, 1981
A>Title: Nucleotide sequence of Moloney murine leukaemia virus.
A:Reference number: A93265; MUID:82035843; PMID:6169994

```

```

A:Accession: A039930
A:Molecule type: genomic RNA
A:Residues: 1-538 <SHt>
A:Experimental source: clone pMLV-1
C:Comment: This protein is synthesized as a gag-pol polyprotein.
C:Gene(s):
A:Gene: gag
C:Superfamily: mammalian retrovirus gag polyprotein I
C:Keywords: core protein, inner coat protein, nucleoprotein, polyprotein
F:2-131/Product: core protein p15 #status predicted <Cl5>
F:132-215/Product: inner coat protein p12 #status predicted <Cl2>
F:216-478/Product: core shell protein p30 #status predicted <Cl3>
F:479-534/Product: nucleoprotein p10 #status predicted <Cl0>

Query Match          7.3%; Score 92; DB 1; Length 538;
Best Local Similarity 26.5%; Pred. No. 1.7;
Matches    45; Conservative   12; Mismatches   45; Indels   68; Gaps    9;

OY      21 RATPAGGGHY-----LHTRPAP---SDN-LPLVLI-EHPPIRDPKKNLTLSL 69
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB     122 RSTPRSSLYPALTPSLGAKPKQVLSDSGEPDLDLTLEDPPPYDPK----- 169

OY      70 LGGHNDPGFMATSPEDRPGGGGA--AGSAEDLAELDLRLOR-----PSGAMPS 118
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB     170 -----PPSDNDNGGEATPAGEADPDSPMASRLRGREPEPVADSTTSQAEP 217

OY      119 EIKLEFESEGLAQKKQRSLSKKLRRKLMWMSQTFCVLVYANDLSRF 168
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB     218 RAGG-----NGOLQYWPFS--DYVMKNNNSF 245

RESULT 5
A97471
hypothetical protein AGR_C_1672 [imported] - Agrobacterium tumefaciens (strain C58, Cer
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: A97471
R:Goodner, B.; Hinkle, G.; Gatungu, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A:Reference number: A97359; PMID:11743194
A:Accession: A97471
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <KUD>
A:Cross-references: CB:AEO07869; PIDN:AAK86722.1; PUD:g15155912; GSDB:GN00169
C:Gene(s):
A:Gene: AGR_C_1672
A:Map position: circular chromosome

Query Match          6.9%; Score 88; DB 2; Length 278;
Best Local Similarity 27.9%; Pred. No. 1.9;
Matches    43; Conservative   19; Mismatches   74; Indels   18; Gaps    6;

OY      16 VVLGRATPAGQHLYHTRAPSDNLPLVDLIEHDDPIFDPERK-DLNETTLRLSLGGHY 74
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB     104 VVVEIDFKOPARRNLNTAPILTS----LSRLKERDGSAFPEDALAGERLAADPHRHLL 159

OY      75 DPGFMATSPEDRPGGGGAAGAEADLAELDOLLR---QRSGAMPSTIKGLE-----FS 126
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB     160 NPTVTATEPR-IASRGAGEGAGDALDEAAMAARPSRAADANGPELSVAIDICFE 218

OY      127 EGLAQKKORL-----SKTLRRKLQMWLWSOTFCP 156
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB     219 KLETVERERLMPARSAKLLLRALLSLAHYAP 252

RESULT 6
AF2689
conserved hypothetical protein Atu0918 [imported] - Agrobacterium tumefaciens (strain C
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
```

[illegible]

QY 3 RCPSLGVTALVAVVG---LRA-----TPAGGHYLIHRAPSDNLPLVDLIEHPDI 53
 DB 235 RAALGAVVALGATGPPLLRAPPPSPPPG-----ADPGS-----BSP 277
 QY 54 FDPKEDLNFTLLSLGCHYD-----PGFMATSPPEDRP-----GGGGAAGAEDL--- 101
 DB 278 -----DPPALIVKLLQEHLEBEVAPRALPPKPKAPATYLANGGSPPSLDQDAEW 331
 QY 102 -----AELDOLIRORPSGAM-----PSEIKGLEFSEGLAOGKKORLSKLRKLOMMLW 150
 DB 332 WGDISEEVNEKLRDPDGTFLVRDASSKIQG-EYTLTLRKGNKKLIXVHRD-GHYGF 389
 QY 151 SQ---TFCPVL 158
 DB 390 SEPLTFCSVV 399

RESULT 10

C75266
 UDP-N-acetylmuramoylalanine-D-glutamate ligase - Deinococcus radiodurans. (strain R1)
 C/Species: Deinococcus radiodurans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C/Accession: C75266
 R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A/Reference number: A75250; MUID:20036896; PMID:10567266
 A/Accession: C75266
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-457 <WHI>
 A/Cross-references: GB:AE002079; GB:AE000513; NID:96460315; PIDN:AAF12038.1; PID:9646031
 A/Experimental source: strain R1
 C/Genetics:
 A/Map position: 1
 C/Suprafamily: UDP-N-acetylmuramate-alanine ligase

Query Match 6.8%; Score 86.5; DB 2; Length 457;
 Best Local Similarity 28.3%; Pred. No. 4.6;
 Matches 34; Conservative 8; Mismatches 39; Indels 39; Gaps 3;

QY 9 VTLVALVVLGLRATPAGGCHYH-----IRPAPSDNPLVDLIEHPDIFDPKE 58
 DB 21 VNLGGRVLIYGLGRSGRVAFPLHGEVSAPFMDLRAPED----- 61
 QY 59 KDLNETLLRL-----LGHYDPCFMATSPPEDRPGGGGAAGAEDLAELOQLROR 112
 DB 62 ---EALMRQLGHRQADLGGTYDLVVAAPGVPIDHRLRLVLAGKGAELIGVALARLR 117

RESULT 11

T45134
 hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)
 C/Species: Microbacterium ammoniaphilum
 C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C/Accession: T45134
 R.Striebel, H.M.; Seiber, S.; Jarsch, M.; Kessler, C.
 Gene 172, 41-46, 1996
 A>Title: Cloning and characterization of the Mami restriction-modification system from M
 A/Reference number: Z22923; MUID:96257250; PMID:8654988
 A/Accession: T45134
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-529 <STR>
 A/Cross-references: EMBL:X79027; NID:9984667; PIDN:CAA55649.1; PID:G1679831
 A/Experimental source: ATCC 15354

Query Match 6.8%; Score 86.5; DB 2; Length 529;
 Best Local Similarity 23.9%; Pred. No. 5.4;

Matches 47; Conservative 18; Mismatches 73; Indels 59; Gaps 8;
 QY 19 GLATPAGGCHYHIRPA--PSNPLVDLIEHPDIFDPKEDLNFTLLSLGCHYD 76
 DB 321 GVGARPARPRRLHPRADLPHGGQGVARLDHPQLEAGASRRPRRLRL--HQR 378
 QY 77 GFMTSPPEDR-----PGGGGAAGAEDLAELOQLRQ--RPSGAMPSEIKGLEFSEGL 129
 DB 379 GDLIRRRRDLRGRRPGQGAHGG-----QGLRRAGRRRGVPSQLAVLR--QGV 428
 QY 130 AGGKKORLSKLRKLOMMLWSQTCFVLYAMNDLGSREMPRYVKGSCFSKSCSVPEG 189
 DB 429 CQAS-----ATLALMTSGRLREL-----AG 450
 QY 190 MVCKPSKSVHLYLWR 206
 DB 451 LLCRDVSYESHLDVTWR 467

RESULT 12

G96762
 hypothetical protein FED5.2 [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C/Accession: G96762
 R.Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzilli
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: G96762
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-622 <STO>
 A/Cross-references: GB:AE005173; NID:91092369; PIDN:AA012777.1; GSPDB:GN00141
 C/Genetics:
 A/Map position: 1
 A/Map position: 1

Query Match 6.8%; Score 86.5; DB 2; Length 622;
 Best Local Similarity 23.2%; Pred. No. 6.5;
 Matches 48; Conservative 21; Mismatches 85; Indels 53; Gaps 9;

QY 19 GLATPAGGCHYHIRPA--PSNPLVDLIEHPDIFDPKEDLNFTLLSLGCHYD 72
 DB 223 GLSATP-----RPSNLTAEIYLSQSSNP--TPGSSFNHTDYSWMASSGGGNS 271
 QY 73 HYDPG--FMATSPPEDR-----GGGGAAGAEDLAE 103
 DB 272 NFGPGAVFGSKCPTRPRNVEEDGPAKPTAAGTAGAGRRHYSGSGGGGCHYAP 331
 QY 104 LDQLLRORBSGAMPSEIKGLEFSEGLAOGKKORLSKLRKLOMMLWSQTCFVLYAND 163
 DB 332 NPMFSPNTGGGGGTAAG--NAPVGGKRQDGN---GRDLHFMVWSSASPVSDVGG 385
 QY 164 LGSRFWRPRYVKGSCFSKR--SCSVPEG 189
 DB 386 GGGNHADYSTATNDHQDKVKTISVPG 412

RESULT 13

T02033
 calcium/calmodulin-dependent protein kinase homolog - maize
 C/Species: Zea mays (maize)
 C/Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 17-Nov-2000
 C/Accession: T02033; T02994; T01694
 R.Lu, Y.T.; Hidaka, H.; Feldman, L.J.

Planta 199, 18-24, 1996

A;Title: Characterization of a calcium/calmodulin-dependent protein kinase homolog from A;Reference number: Z14504; MUID:96236830; PMID:8680305

A;Accession: T02033

A;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A;Residues: 1-625 <LU>

A;Cross-references: EMBL:S82324; NID:g1839596; PIDN:AAB47181.1; PID:g1839597

A;Experimental source: cv. Merit

R;Furumoto, T.; Ogawa, N.; Hata, S.; Izui, K.

PBS Lett. 396, 147-151, 1996

A;Title: Plant calcium-dependent protein kinase-related kinases (CRK) do not require cal A;Reference number: Z14398; MUID:97072168; PMID:8914977

A;Accession: T02994

A;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A;Residues: 29-146, 'GA', 147-625 <FUR>

A;Cross-references: EMBL:D84507; NID:g1313906; PIDN:BAA12691.1; PID:g1313907

A;Experimental source: strain inbred line H84; root

A;Accession: T01694

A;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A;Residues: 174-398, 'G', 400-625 <FU2>

A;Cross-references: EMBL:D38452; NID:g2443387; PIDN:BAA22410.1; PID:g2443388

A;Experimental source: leaf

C;Genetics:

A;Note: BAA12691.1

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo F;172-436/Domain: protein kinase homology <KIN>

Query Match 6.8%; Score 86; DB 2; Length 625;
Best Local Similarity 32.3%; Pred.No. 7.3;
Matches 42; Conservative 12; Mismatches 46; Indels 30; Gaps 8;

QY 24 PAGOHYL--HIRPAPSDNLPLVDLIHPPDIPDPKSKDLNETLLRSILGGHYDPPGMAT 81
| | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 73 PAGASPLPAGVSPSPARSTPR-RFFKRPPPPSP-AKHIKATLAKLGGGPKPEG---T 127
| | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 82 SPPEPD-----RPGGGGAAGAAGDAELDOLLQRPSGAMPSEIKGLEFSEGIAQ--GK 133
| | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 128 IPEGGAGAGAGAGAGAGAAGVGAADSAEAD-----RP-----LDKTFFGAKNFGA 172
| | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 134 QORLSKKLR 143
| | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 173 KYDLGKEVR 182
| | | | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 14

B38749

3-phosphatidylinositol kinase (EC 2.7.1.-) 85K chain B - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 16-Jul-1999

C;Accession: B38749

R;Otsu, M.; Hiles, I.; Gout, I.; Fry, M.J.; Ruiz-Larrea, F.; Panayotou, G.; Thompson, A.

Cell 65, 91-104, 1991

A;Title: Characterization of two 85 kd proteins that associate with receptor tyrosine ki A;Reference number: A38749; MUID:91191567; PMID:1707345

A;Accession: B38749

A;Status: preliminary

A:Molecule type: mRNA

A;Residues: 1-723 <OTS>

A;Cross-references: GB:M61745; GB:M61746

C;Superfamily: SH2 homology

C;Keywords: phosphotransferase

F;325-420/Domain: SH2 homology <SH2A>

F;617-706/Domain: SH2 homology <SH2>

Query Match 6.8%; Score 86; DB 2; Length 723;
Best Local Similarity 25.9%; Pred.No. 8.6;
Matches 49; Conservative 23; Mismatches 59; Indels 58; Gaps 12;

QY 3 RCPSLGVTLXALVVVLG---LRA-----TPAGGHYLHIRPAPSDNLPLVDLIHPDPPIF 54

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2003, 09:19:22 ; Search time 37 Seconds
(without alignments)
835.517 Million cell updates/sec

Title: US-09-897-322-2

Perfect score: 1268

Sequence: 1 MERCPSLGVTLYALVVLGL.....QRGWIPIQYPIISECKKC 232

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq 101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1268	100.0	232	20 AAW96279	Spemann organiser
2	1268	100.0	232	23 AAG79341	Human Noggin. Hom
3	1261	99.4	232	15 AAR50303	Human noggin. Hom
4	1258	99.2	232	15 AAR50304	Murine noggin. Mu
5	1258	99.2	232	20 AAW96280	Spemann organiser
6	1258	99.2	232	23 AAG79347	Partial mouse Nogg
7	1195	94.2	449	20 AAW96278	Human noggin/immun
8	112	8.8	361	21 AAB32696	Eucalyptus grandis
9	102	8.0	2392	21 AAB07665	Amino acid sequenc
10	99.5	7.8	413	21 AAY53820	Amino acid sequenc

11	99	7.8	836	22	AA85703	Novel protein kina
12	93.5	7.4	372	18	AAW22065	Chick fringe A (ra
13	92.5	7.3	319	18	AAW18658	Fragmented human N
14	92	7.3	538	21	AAI10043	WMLV gag protein.
15	92	7.3	648	20	AAV17946	WMLV gag gene pro
16	92	7.3	761	22	ABG02677	Novel human diagno
17	92	7.3	1737	21	AAI10044	WMLV gag-pol prote
18	91	7.2	743	22	AAW79738	Human protein SEQ
19	90.5	7.1	623	22	AAW85583	Rice CDPK (clone r
20	89	7.0	1093	22	AAE02455	Mouse semaphorin G
21	88.5	7.0	522	19	AAW69341	Secreted protein o
22	88.5	7.0	623	17	AAW03140	Human IgG CHI-hing
23	88.5	7.0	979	19	AAW40309	Human ITAK protein
24	88.5	7.0	979	21	AAV95294	Human GEF containi
25	87.5	6.9	235	22	ABB61683	Drosophila melanog
26	87.5	6.9	728	22	AAW78754	Human protein SEQ
27	87.5	6.9	1368	23	AAU77182	Human kinesin moto
28	87	6.9	14	15	AAW49838	Noggin conserved p
29	87	6.9	14	15	AAW50305	Noggin conserved r
30	87	6.9	14	20	AAW96281	Conserved sequence
31	87	6.9	14	23	AAG79342	Noggin conserved p
32	87	6.9	1975	22	ABB62094	Drosophila melanog
33	86.5	6.8	622	21	AAW26950	Auxin transport pr
34	86.5	6.8	622	23	ABB91576	Herbicidally activ
35	86.5	6.8	939	23	AAG68296	Human semaphorin G
36	86.5	6.8	954	23	AAG68295	Human semaphorin G
37	86.5	6.8	999	21	AAV94990	Human secreted pro
38	86.5	6.8	1034	23	AAG68291	Human semaphorin G
39	86.5	6.8	1049	23	AAG68289	Human semaphorin G
40	86.5	6.8	1078	23	AAG68292	Human semaphorin G
41	86.5	6.8	1092	23	ABG34077	Human Pro peptide
42	86.5	6.8	1093	23	AAG68290	Human semaphorin G
43	86.5	6.8	1136	23	AAG68294	Human semaphorin G
44	86.5	6.8	1151	23	AAG68293	Human semaphorin G
45	86.5	6.8	1352	23	AAE18212	Human MOL4 protein

ALIGNMENTS

RESULT 1

AAW96279

ID AAW96279 standard; Protein; 232 AA.

XX
AC AAW96279;

XX
DT 14-JUN-1999 (first entry)

XX
DE Spemann organiser signal Noggin polypeptide.

XX
KW Spemann organiser signal; noggin; bone morphogenetic protein; BMP;
KW disease; disorder; bone; bone growth; trauma; burns;
KW Fibrodysplasia Ossificans Progressiva; FOP; antibodies; detection;
KW diagnosis; therapy.

OS Homo sapiens.

XX
PN WO9903996-A1.

XX
PD 28-JAN-1999.

XX
PF 17-JUL-1999; 98WO-US14603.

XX
PR 17-JUL-1997; 97US-0897236.

XX
PA (REGC-) REGENERON PHARM INC.

XX
PA (REGC) UNIV CALIFORNIA.

XX
PI Economides AN, Harland RM, Stahl N;

XX
DR WPI; 1999-132240/11.

XX
DR N-PSDB; AAX09017.

```

PT New modified Noggin polypeptides - having an amino acid deletion to
PT increase bioavailability in animal serum, used for treating bone
PT disorders or diseases
XX
XX Claim 1; Figure 1, 132pp; English.
XX
CC Noggin polypeptides induce dorsal growth and block bone
CC morphogenetic protein (BMP) activity in vertebrates. Modified Noggin
CC polypeptides, modified by a deletion of amino acids 138-144, can be
CC used for the treatment of a disease or disorder affecting bone,
CC e.g. abnormal bone growth following hip replacement surgery, trauma,
CC burns, or spinal cord injury, or fibrodysplasia Ossificans
CC progressive (FOP). This is achieved by the modified Noggin
CC polypeptide exhibiting improved bioavailability in animal
CC sera while retaining the ability to bind to a BMP. The products can
CC also be used for the production of antibodies, detection and
CC diagnosis. The modified Noggin protein (AAW96278) has enhanced
CC therapeutic properties.
XX
SQ Sequence 232 AA;
Query Match 100.0%; Score 1268; DB 20; Length 232;
Best Local Similarity 100.0%; Pred. No. 5e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MERCPSELGVTLYLVVVLGIRATPAGGQHYLHIRPASPNDLPLVDLIEHPDPIFPKEXD 60
DB 1 MERCPSELGVTLYLVVVLGIRATPAGGQHYLHIRPASPNDLPLVDLIEHPDPIFPKEXD 60
OY 61 LNETLRLSLGHHYDGFMAFSPEDRPGGGGAGAGAEDELAEIDLRLRORSGAMPSEI 120
DB 61 LNETLRLSLGHHYDGFMAFSPEDRPGGGGAGAGAEDELAEIDLRLRORSGAMPSEI 120
OY 121 KGLEFSEGLAOGKKQRLSKRLRRKLQMWLMSQTFPCVLYAMNDLGSFPMRYVVGSCFS 180
DB 121 KGLEFSEGLAOGKKQRLSKRLRRKLQMWLMSQTFPCVLYAMNDLGSFPMRYVVGSCFS 180
OY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCQRRGQRCGMIPQYPIISECKSC 232
DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCQRRGQRCGMIPQYPIISECKSC 232
RESULT 2
AAG79341
ID AAG79341 standard; Protein; 232 AA.
XX
AC AAG79341;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human Noggin.
XX
KW Human; noggin; neurotrophic; growth factor; dorsal development;
KW vertebrate; fibroblast growth factor; FGF; cognate receptor; cancer;
KW Kunitz-type protease inhibitor; nerve; muscle; bone; neurodegeneration;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; peripheral neuropathy; culture media;
KW traumatic nerve injury; diabetes; kidney dysfunction; anencephaly.
XX
XX Homo sapiens.
XX
XX US6277593-B1.
XX
XX 21-AUG-2001.
XX
PD 21-AUG-2001.
XX
PF 07-OCT-1998; 98US-0167874.
XX
XX 02-SEP-1993; 93US-0392935.
XX 02-SEP-1993; 93WO-US08326.
XX 07-JUN-1995; 95US-0485721.
XX 03-SEP-1992; 92US-0939954.
XX 23-SEP-1992; 92US-0950410.
XX 06-OCT-1992; 92US-0957401.

```

```

XX (REG-) REGENERON PHARM INC.
PA (REG) UNIV CALIFORNIA.
PA
XX Valenzuela DM, Ip NY, Cudny HD, Yancopoulos GD, Harland RM;
XX Smith WC, Lamb T, Knecht A;
XX MPI: 2002-442065/47.
XX N-PSDB; AAI72927.
DR
PT Recombinant preparation of noggin polypeptide, useful e.g. for treating
PT neurodegenerative disease, by expressing specific human sequence
XX
XX Example 4; Fig 1; 40pp; English.
PS
CC This sequence represents human Noggin polypeptide. Noggin is a
CC neurotrophic growth factor which induces dorsal development in
CC vertebrates. Noggin modifies the actions of fibroblast growth factor
CC (FGF). It is a 26 kD protein with a hydrophobic amino terminal. Noggin
CC is secreted, apparently as a dimeric glycoprotein. The carboxy terminal
CC region of Noggin shows homology to a Kunitz-type protease inhibitor.
CC Noggin polypeptide may be prepared by culturing cells transformed with
CC a vector that contains a control sequence operatively linked to a
CC nucleic acid molecule which comprises the coding region for human
CC noggin or a sequence encoding the same amino acid sequence. Human
CC Noggin, also its fusion proteins and derivatives, may be used to raise
CC specific antibodies (Ab), for diagnosis, for detection and purification
CC of Ab, to induce growth of nerve and muscle cells in mammals, and to
CC regulate bone or muscle growth, e.g. in wound-healing compositions and
CC for treating neurodegeneration (Alzheimer's, Parkinson's or
CC Huntington's diseases, amyotrophic lateral sclerosis and peripheral
CC neuropathy), traumatic nerve injury, diabetes, kidney dysfunction, the
CC toxic effects of chemotherapeutic agents being used to treat acquired
CC immune deficiency syndrome or cancer, and congenital malformations such
CC as anencephaly, as an additive to culture media used for growing nerve
CC cells and to isolate cognate receptors, potentially useful for
CC diagnosis of some cancers. Ab's are used for in vitro or in vivo
CC therapy or diagnosis and for purification of Noggin.
XX
SQ Sequence 232 AA;
Query Match 100.0%; Score 1268; DB 23; Length 232;
Best Local Similarity 100.0%; Pred. No. 5e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MERCPSELGVTLYLVVVLGIRATPAGGQHYLHIRPASPNDLPLVDLIEHPDPIFPKEXD 60
DB 1 MERCPSELGVTLYLVVVLGIRATPAGGQHYLHIRPASPNDLPLVDLIEHPDPIFPKEXD 60
OY 61 LNETLRLSLGHHYDGFMAFSPEDRPGGGGAGAGAEDELAEIDLRLRORSGAMPSEI 120
DB 61 LNETLRLSLGHHYDGFMAFSPEDRPGGGGAGAGAEDELAEIDLRLRORSGAMPSEI 120
OY 121 KGLEFSEGLAOGKKQRLSKRLRRKLQMWLMSQTFPCVLYAMNDLGSFPMRYVVGSCFS 180
DB 121 KGLEFSEGLAOGKKQRLSKRLRRKLQMWLMSQTFPCVLYAMNDLGSFPMRYVVGSCFS 180
OY 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCQRRGQRCGMIPQYPIISECKSC 232
DB 181 KRSCSVPEGWVCKPSKSVHLTVLRMRCQRRGQRCGMIPQYPIISECKSC 232
RESULT 3
AAR50303
ID AAR50303 standard; Protein; 232 AA.
XX
AC AAR50303;
XX
DT 19-OCT-1994 (first entry)
XX
DE Human noggin.
XX
KW Human; noggin; hydrophobic amino terminal; kunitz-type; bone growth;

```


protease inhibitor; regulation; cartilage; growth factor; epidermis;
tissue matrix; potentiation; wound healing; diagnosis; probe; tumour;
fibroblast growth factor; FGF; activin; nerve; muscle cell;
Alzheimers disease; Parkinsons disease; Huntington's chorea;
peripheral neuropathy.

XX Homo sapiens.

XX WO9405791-A.

XX 17-MAR-1994.

XX 02-SEP-1993; 93WO-US08326.

XX 03-SEP-1992; 92US-0939954.

XX 23-SEP-1992; 92US-0950410.

XX 06-OCT-1992; 92US-0957401.

XX (REGE-) REGENERON PHARM INC.

XX Cudny H, Harland RM, Ip NY, Knecht A, Lamb T, Smith WC;

XX Valenzuela DM, Yancopoulos GD;

XX WPI: 1994-101196/12.

XX N-PSDB; AAQ76342.

XX Noggin protein capable of inducing dorsal growth, and sequences
encoding it - useful for treating neurodegenerative disorders and
neural damage, e.g. due to trauma or after chemotherapy

XX Claim 1; Fig 1; 100pp; English.

XX This sequence represents human noggin protein. The noggin cDNA

encodes a 26 kD secreted protein which has a hydrophobic amino

terminal sequence. The carboxy terminal sequence of noggin shows

homology to a kunitz-type protease inhibitor, indicating that it may

exhibit activities of a protease inhibitor. Noggin is a regulator of

cartilage production and a growth factor for tissue matrix and

epidermis. Noggin is useful for regulating cartilage and bone growth,

optionally in conjunction with other growth factors which may be

potentiated by noggin. It is also useful in wound healing and in the

isolation of its receptor, which may itself be used as a diagnostic

probe for certain types of tumour. Noggin modifies the actions of

fibroblast growth factor (FGF) and also activin. Noggin may be used

for enhancing the survival or inducing the growth of nerve and muscle

cells. It may therefore be useful in the therapy of congenital

conditions or degenerative disorders of the nervous system, eg.

Alzheimers disease, Parkinsons disease, Huntington's chorea and or

peripheral neuropathy.

XX Sequence 232 AA;

Query Match 99.4%; Score 1261; DB 15; Length 232;

Best Local Similarity 99.6%; Pred. No. 2.6e-122;

Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDPIFDPEKD 60

Db 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDPIFDPEKD 60

QY 61 LNETLRLSLGGHYDPGFMTATPPEDRPGGGGAAGAEDELAELOLLRORPSCAMPSEI 120

Db 61 LNETLRLSLGGHYDPGFMTATPPEDRPGGGGAAGAEDELAELOLLRORPSCAMPSEI 120

QY 121 KGLEFSEGLAQGKKRLSKLRKLQWLWSQTFPCVLYAWNDLGRFRPYYKVGSCFS 180

Db 121 KGLEFSEGLAQGKKRLSKLRKLQWLWSQTFPCVLYAWNDLGRFRPYYKVGSCFS 180

QY 181 KRSCSVPEGMVCKPSKSVHLTVLRWRCORRGQRCGWIPIQYPIISECKSC 232

Db 181 KRSCSVPEGMVCKPSKSVHLTVLRWRCORRGQRCGWIPIQYPIISECKSC 232

RESULT 4

AAR50304

ID AAR50304 standard; Protein; 232 AA.

XX AAR50304;

XX 19-OCT-1994 (first entry)

XX Murine noggin.

XX Human; noggin; hydrophobic amino terminal; kunitz-type; bone growth;

protease inhibitor; regulation; cartilage; growth factor; epidermis;

tissue matrix; potentiation; wound healing; diagnosis; probe; tumour;

fibroblast growth factor; FGF; activin; nerve; muscle cell;

Alzheimers disease; Parkinsons disease; Huntington's chorea;

peripheral neuropathy.

XX Mus musculus.

XX WO9405791-A.

XX 17-MAR-1994.

XX 02-SEP-1993; 93WO-US08326.

XX 03-SEP-1992; 92US-0939954.

XX 23-SEP-1992; 92US-0950410.

XX 06-OCT-1992; 92US-0957401.

XX (REGE-) REGENERON PHARM INC.

XX Cudny H, Harland RM, Ip NY, Knecht A, Lamb T, Smith WC;

XX Valenzuela DM, Yancopoulos GD;

XX WPI: 1994-101196/12.

XX N-PSDB; AAQ44784.

XX Noggin protein capable of inducing dorsal growth, and sequences

encoding it - useful for treating neurodegenerative disorders and

neural damage, e.g. due to trauma or after chemotherapy

XX Disclosure; Fig 13; 100pp; English.

XX This sequence represents murine noggin protein. The noggin cDNA

encodes a 26 kD secreted protein which has a hydrophobic amino

terminal sequence. The carboxy terminal sequence of noggin shows

homology to a kunitz-type protease inhibitor, indicating that it may

exhibit activities of a protease inhibitor. Noggin is a regulator of

cartilage production and a growth factor for tissue matrix and

epidermis. Noggin is useful for regulating cartilage and bone

growth, optionally in conjunction with other growth factors which

may be potentiated by noggin. It is also useful in wound healing

and in the isolation of its receptor, which may itself be used as

a diagnostic probe for certain types of tumour. Noggin modifies the

actions of fibroblast growth factor (FGF) and also activin. Noggin

may be used for enhancing the survival or inducing the growth of

nerve and muscle cells. It may therefore be useful in the therapy

of congenital conditions or degenerative disorders of the nervous

system, eg. Alzheimers disease, Parkinsons disease, Huntington's

chorea and or peripheral neuropathy.

XX Sequence 232 AA;

Query Match 99.2%; Score 1258; DB 15; Length 232;

Best Local Similarity 99.1%; Pred. No. 5.4e-122;

Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDPIFDPEKD 60

Db 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDPIFDPEKD 60

QY 61 LNETLRLSLGGHYDPGFMTATPPEDRPGGGGAAGAEDELAELOLLRORPSCAMPSEI 120

Db 61 INETLLRSILGHHYDGFMAATSPEDRPGGGGAGAGADLAELDQLLRQSPGAMPESEI 120
 QY 121 KGLEFSEGLAOGKKORLSKTLRRKLQMWLMSQTFCEPVLVYAMNDLGSFMPRYVYVSGCFS 180
 Db 121 KGLEFSEGLAOGKKORLSKTLRRKLQMWLMSQTFCEPVLVYAMNDLGSFMPRYVYVSGCFS 180
 QY 181 KRSCSVPEGMVCKPSKSVHLTVLRMRCORRGRCGMIPIQYPIISECKCSC 232
 Db 181 KRSCSVPEGMVCKPSKSVHLTVLRMRCORRGRCGMIPIQYPIISECKCSC 232

RESULT 5
 AAM96280
 ID AAM96280 standard; protein; 232 AA.
 AC AAM96280;
 DT 14-JUN-1999 (first entry)
 DE Spemann organiser signal Noggin polypeptide.
 XX
 XX Spemann organiser signal; noggin; bone morphogenetic protein; BMP;
 KW disease; disorder; bone; bone growth; trauma; burns;
 KW Fibrodysplasia Ossificans Progressiva; FOP; antibodies; detection;
 KW diagnosis; therapy.
 XX
 OS Mus musculus.
 PN WO903996-A1.
 XX
 XX 28-JAN-1999.
 PD
 PF 17-JUL-1998; 98WO-US14603.
 XX
 XX 17-JUL-1997; 97US-0897236.
 PR
 XX (REGC-) REGENERON PHARM INC.
 PA (REGC) UNIV CALIFORNIA.
 PI Economides AN, Harland RM, Stahl N;
 XX
 XX WPI; 1999-132240/11.
 DR N-PSDB; AAX09018.
 XX
 XX New modified Noggin polypeptides - having an amino acid deletion to
 PT increase bioavailability in animal serum, used for treating bone
 PT disorders or diseases
 PS Disclosure; Figure 13; 132pp; English.
 XX
 XX Noggin polypeptides induce dorsal growth and block bone
 CC morphogenetic protein (BMP) activity in vertebrates. Modified Noggin
 CC polypeptides, modified by a deletion of amino acids 138-144, can be
 CC used for the treatment of a disease or disorder affecting bone.
 CC e.g. abnormal bone growth following hip replacement surgery, trauma,
 CC burns, or spinal cord injury, or Fibrodysplasia Ossificans
 CC Progressiva (FOP). This is achieved by the modified Noggin
 CC polypeptide exhibiting improved bioavailability in animal
 CC sera while retaining the ability to bind to a BMP. The products can
 CC also be used for the production of antibodies, detection and
 CC diagnosis. The modified Noggin protein has enhanced therapeutic
 CC properties.
 XX
 XX Sequence 232 AA;
 SQ

Query Match 99.2%; Score 1258; DB 20; Length 232;
 Best Local Similarity 99.1%; Pred. No. 5, 4e-122;
 Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MERCPSELGVTLVVVGLRATPAGGQHYLHRPAPSDNLPLVDLIEHDPPIFDPKXKD 60
 Db 1 MERCPSELGVTLVVVGLRATPAGGQHYLHRPAPSDNLPLVDLIEHDPPIFDPKXKD 60

QY 61 INETLLRSILGHHYDGFMAATSPEDRPGGGGAGAGADLAELDQLLRQSPGAMPESEI 120
 Db 61 INETLLRSILGHHYDGFMAATSPEDRPGGGGAGAGADLAELDQLLRQSPGAMPESEI 120
 QY 121 KGLEFSEGLAOGKKORLSKTLRRKLQMWLMSQTFCEPVLVYAMNDLGSFMPRYVYVSGCFS 180
 Db 121 KGLEFSEGLAOGKKORLSKTLRRKLQMWLMSQTFCEPVLVYAMNDLGSFMPRYVYVSGCFS 180
 QY 181 KRSCSVPEGMVCKPSKSVHLTVLRMRCORRGRCGMIPIQYPIISECKCSC 232
 Db 181 KRSCSVPEGMVCKPSKSVHLTVLRMRCORRGRCGMIPIQYPIISECKCSC 232

RESULT 6
 AAG79347
 ID AAG79347 standard; protein; 232 AA.
 AC AAG79347;
 DT 21-AUG-2002 (first entry)
 DE Partial mouse Noggin.
 XX
 XX Human; noggin; neurotrophic; growth factor; dorsal development;
 KW vertebrate; fibroblast growth factor; FGF; cognate receptor; cancer;
 KW Kunitz-type protease inhibitor; nerve; muscle; bone; neurodegeneration;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; peripheral neuropathy; culture media;
 KW traumatic nerve injury; diabetes; kidney dysfunction; anencephaly.
 XX
 OS Mus musculus.
 PN US6277593-B1.
 XX
 XX 21-AUG-2001.
 PD
 PF 07-OCT-1998; 98US-0167874.
 XX
 XX 02-SEP-1993; 93US-0392935.
 PR 02-SEP-1993; 93WO-US08326.
 PR 07-JUN-1995; 95US-0485721.
 PR 03-SEP-1992; 92US-0939954.
 PR 23-SEP-1992; 92US-0950410.
 PR 06-OCT-1992; 92US-0957401.
 XX
 XX (REGC-) REGENERON PHARM INC.
 PA (REGC) UNIV CALIFORNIA.
 PI Valenzuela DM, Ip NY, Cudny HD, Yancopoulos GD, Harland RM;
 PI Smith WC, Lamb T, Knecht A;
 XX
 XX WPI; 2002-442065/47.
 DR N-PSDB; AAI72930.
 XX
 XX Recombinant preparation of noggin polypeptide, useful e.g. for treating
 PT neurodegenerative disease, by expressing specific human sequence -
 PT
 PS Example 3; Fig 13; 40pp; English.
 XX
 XX This sequence represents partial mouse Noggin polypeptide. Noggin is
 CC a neurotrophic growth factor which induces dorsal development in
 CC vertebrates. Noggin modifies the actions of fibroblast growth factor
 CC (FGF). It is a 26 kD protein with a hydrophobic amino terminal. Noggin
 CC is secreted, apparently as a dimeric glycoprotein. The carboxy terminal
 CC region of Noggin shows homology to a Kunitz-type protease inhibitor.
 CC Noggin polypeptide may be prepared by culturing cells transformed with
 CC a vector that contains a control sequence operatively linked to a
 CC nucleic acid molecule which comprises the coding region for human
 CC noggin or a sequence encoding the same amino acid sequence. Human
 CC Noggin, also its fusion proteins and derivatives, may be used to raise
 CC specific antibodies (Ab), for diagnosis, for detection and purification
 CC of Ab, to induce growth of nerve and muscle cells in mammals, and to
 CC regulate bone or muscle growth, e.g. in wound-healing compositions and

CC for treating neurodegeneration (Alzheimer's, Parkinson's or
CC Huntington's diseases, amyotrophic lateral sclerosis and peripheral
CC neuropathy), traumatic nerve injury, diabetes, kidney dysfunction, the
CC toxic effects of chemotherapeutic agents being used to treat acquired
CC immune deficiency syndrome or cancer, and congenital malformations such
CC as anencephaly, as an additive to culture media used for growing nerve
CC cells and to isolate cognate receptors, potentially useful for
CC diagnosis of some cancers. Ab's are used for in vitro or in vivo
CC therapy or diagnosis and for purification of Noggin.
XX
SQ Sequence 232 AA;

Query Match 99.2%; Score 1258; DB 23; Length 232;
Best Local Similarity 99.1%; Pred. No. 5.4e-122;
Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDIFDPKEKD 60
DB 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDIFDPKEKD 60
QY 61 LNETLLRSLGGHYDPGFMTATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120
DB 61 LNETLLRSLGGHYDPGFMTATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120
QY 121 KGLEFSEGLAQKKQRLSKLRKQLQMWLWSQTFCPVLYAMNDLGSRFWRPYYKVGSCFS 180
DB 121 KGLEFSEGLAQKKQRLSKLRKQLQMWLWSQTFCPVLYAMNDLGSRFWRPYYKVGSCFS 180
QY 181 KRCSVPGEWCKPSKSVHLTVLRWRCRRGQRCGWIPQIPIIIECKKSC 232
DB 181 KRCSVPGEWCKPSKSVHLTVLRWRCRRGQRCGWIPQIPIIIECKKSC 232

RESULT 7
AAW96278
ID AAW96278 standard; Protein; 449 AA.

XX AAW96278;
XX
XX
DT 14-JUN-1999 (first entry)
DE Human noggin/immunoglobulin G1 fusion peptide (hNGdelta133-144Fc).
DE
KW Spemann organizer signal; noggin; bone morphogenetic protein; BMP;
KW disease; disorder; bone; bone growth; trauma; burns;
KW Fibrodysplasia Ossificans Progressiva; FOP; antibodies; detection;
KW diagnosis; therapy.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT Peptide 1..30
FT /label= Putative_signal_peptide
FT Modified-site 62
FT /label= Glycosylation_site
FT Modified-site 133..134
FT /note= "Marks position of 133-144 deletion"
FT Disulfide-bond 218
FT /note= "Cysteine residue used in inter-chain
FT disulfide bridge in human noggin"
FT Region 221..222
FT /note= "Ser-Gly bridge connecting human noggin
FT sequence to human IgG1 Fc region"
FT Disulfide-bond 228
FT /note= "Cysteine involved in inter-chain
FT disulfide bridge of IgG hinge preceding
FT the Fc region"
FT Disulfide-bond 232
FT /note= "Cysteine involved in inter-chain
FT disulfide bridge of IgG hinge preceding
FT the Fc region"
FT Domain 223..449
FT /label= Fc_domain

FT Modified-site 299
FT /label= Glycosylation_site
XX W09503996-A1.
XX
XX 28-JAN-1999.
XX
XX 17-JUL-1998; 98WO-US14603.
XX
XX 17-JUL-1997; 97US-0897236.
XX
XX (REGC-) REGENERON PHARM INC.
XX (REGC) UNIV CALIFORNIA.
XX
XX Economides AN, Harland RM, Stahl N;
XX
XX WPI; 1999-132240/11.
XX
XX New modified Noggin polypeptides - having an amino acid deletion to
XX increase bioavailability in animal serum, used for treating bone
XX disorders or diseases
XX
XX Example 9; Figure 14; 132pp; English.
XX
XX Noggin polypeptides induce dorsal growth and block bone
XX morphogenetic protein (BMP) activity in vertebrates. Modified Noggin
XX polypeptides, modified by a deletion of amino acids 138-144, can be
XX used for the treatment of a disease or disorder effecting bone.
XX e.g. abnormal bone growth following hip replacement surgery, trauma,
XX burns, or spinal cord injury, or Fibrodysplasia Ossificans
XX Progressiva (FOP). This is achieved by the modified Noggin
XX polypeptide exhibiting improved bioavailability in animal
XX sera while retaining the ability to bind to a BMP. The products can
XX also be used for the production of antibodies, detection and
XX diagnosis. The modified Noggin protein (AAW96278) has enhanced
XX therapeutic properties.
XX
SQ Sequence 449 AA;

Query Match 94.2%; Score 1195; DB 20; Length 449;
Best Local Similarity 94.8%; Pred. No. 4.5e-115;
Matches 220; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDIFDPKEKD 60
DB 1 MERCPSLGVTLYALVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDIFDPKEKD 60
QY 61 LNETLLRSLGGHYDPGFMTATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120
DB 61 LNETLLRSLGGHYDPGFMTATSPEDRPGGGGAAGGAEDLAELDQLLRQPSGAMPSEI 120
QY 121 KGLEFSEGLAQKKQRLSKLRKQLQMWLWSQTFCPVLYAMNDLGSRFWRPYYKVGSCFS 180
DB 121 KGLEFSEGLAQKKQRLSKLRKQLQMWLWSQTFCPVLYAMNDLGSRFWRPYYKVGSCFS 180
QY 181 KRCSVPGEWCKPSKSVHLTVLRWRCRRGQRCGWIPQIPIIIECKKSC 232
DB 169 KRCSVPGEWCKPSKSVHLTVLRWRCRRGQRCGWIPQIPIIIECKKSC 220

RESULT 8
AAB32696
ID AAB32696 standard; Protein; 361 AA.
XX
XX AAB32696;
XX
XX 25-JAN-2001 (first entry)
XX
XX Eucalyptus grandis transcription factor protein sequence #154.
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;

DE XX Amino acid sequence of the human autosomal highly conserved protein.

KW Human; AHP gene; autosomal highly conserved protein; schizophrenia; neurological disease; genetic predisposition; chromosome 6p23; D6S274; D6S285; psychological disease; gene therapy.

KW D6S285; psychological disease; gene therapy.

OS Homo sapiens.

XX WO9957316-A1.

PN 11-NOV-1999.

XX 30-APR-1999; 99WO-1B00846.

PF 30-APR-1998; 98US-0083625.

PR 31-DEC-1998; 98US-0114592.

XX (INSP) INST PASTEUR.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Leroy P, Bourgeron T, McElreavey K, Fellous M, Jamain S;

PI WPI; 2000-086415/07.

DR N-PSDB; AA236755, AA236756, AA236757.

XX New gene encoding autosomal high conserved protein used to diagnose a genetic predisposition to schizophrenia -

PT Claim 15; Page 61-63; 76pp; English.

PS The present sequence represents the human AHP (autosomal highly conserved protein) gene. The AHP gene is linked to a genetic predisposition to schizophrenia. The gene is located on chromosome 6p23, between markers D6S274 and D6S285. Several polymorphisms are found in the AHP gene. Oligonucleotide probes derived from the AHP sequences can be used to screen for patients having a genetic predisposition for a neurological or psychological disease, especially schizophrenia. The invention is used to diagnose a genetic predisposition to schizophrenia, and to treat the disorder by gene therapy. The invention provides a treatment that is specific to schizophrenia disorders, without the risk of significant side effects.

XX Sequence 413 AA;

SQ Query Match 7.8%; Score 99.5; DB 21; Length 413; Best Local Similarity 27.6%; Pred. No. 0.15; Matches 50; Conservative 17; Mismatches 69; Indels 45; Gaps 11;

QY 26 GGOYHLIRPAPSDNPLVDLIEHP--DPIDPKEDLNELTLRLSLGGHYDPGFMTSP 83

DB 19 GGDH----EPVPSLRGPTTAVPCRPDDPQAEPOA-----PG----- 51

QY 84 PEDRPGGGGAAGAEDELALDQLLRQ-----PSGAMPSEIKLFESEGLA--QKKQRL 137

DB 52 ---RPTAGLAAADKLEPPRELKRGGAAGSGAELOEQAGCAPEAAAPRPARL 108

QY 138 S-KKLRLKQMLWSOTFCFVLVYAWNDLGRFWRPVYK---VGSFCSKRCSPVPEGMVCK 193

DB 109 SAREYSQVHEWLW-QSYCYL-TWHS-GLAAPFAYCSPQSPQSPSGAAVPQAAPP 165

QY 194 P 194

DB 166 P 166

RESULT 11

AA65703

ID AAB65703 standard; Protein; 836 AA.

XX AAB65703;

XX 27-MAR-2001 (first entry)

XX

DE XX Novel protein kinase, SEQ ID NO: 232.

KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic; immunosuppressive; cardiac; renal; antiinflammatory; antiasthmatic; dermatological; antidiabetic; antiinfertility; gene therapy; vaccine; immune disorder; cardiovascular disease; neurodegenerative disease; cancer; autoimmune disorder; stroke; inflammatory bowel disease; inflammatory pelvic disease; multiple sclerosis; psoriasis.

XX Homo sapiens.

XX WO200073469-A2.

PN 07-DEC-2000.

XX 26-MAY-2000; 2000WO-US14842.

PF 28-MAY-1999; 99US-0136503.

PR (SUGE-) SUGEN INC.

XX Plowman GD, Martinez R, Whyte D, Sudersanam S;

PI WPI; 2001-032161/04.

DR N-PSDB; AAF44732.

XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and treating immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers -

PT Claim 10; Fig 1; 310pp; English.

PS The present sequence is a novel protein kinase. The novel protein kinases and the nucleic acids that encode them may be used in the treatment and diagnosis of diseases associated with inappropriate kinase expression such as immune-related diseases and disorders, cardiovascular disease, neurodegenerative diseases and/or cancers. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays. The kinase polypeptides may be used as antigens in the production of antibodies of kinase expression and activity. Anti-kinase antibodies and kinase antagonists may also be used to down regulate kinase expression and activity. Diseases related to kinase expression and activity include rheumatoid arthritis, atherosclerosis, autoimmune disorders, complications of organ transplantation, myocardial infarction, immune disorders, cardiomyopathies, strokes, renal failure, oxidative-stress related disorders, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and reproductive disorders.

XX Sequence 836 AA;

SQ Query Match 7.8%; Score 99; DB 22; Length 836; Best Local Similarity 25.0%; Pred. No. 0.43; Matches 43; Conservative 22; Mismatches 83; Indels 24; Gaps 5;

QY 3 RCPSLGVTLVYLVVLGLRATPAGGQHYLHIRPAPSDNPLVDLIEHPDPIFDPKEKDLN 62

DB 608 KCGQLGVNKKRLGILNLLGGPLGKQVIRVSCGDEFIATD-----EKVLN 655

QY 63 ETLRLSLGG-HYDFGFMTSPPEDRPGGGGAAGAEDELALDQLLRQPSGAMPSEIK 121

DB 656 SNTIRSNSSGLSIGTVFQSSS-----PGGGGGGGGEEEDSQSESTPD-PSGGRGTME 709

QY 122 GLEFSEGL-----AQKKQRLSKLRLKQMLWSQTCPCPVLYAWNDLGRF 168

DB 710 ADRGMEGLISPTTEAMNGNSGASSCCFGWLRKLENAEFIPMDPSPLSAAF 761

RESULT 12

AAW22065

ID AAW22065 standard; Protein; 372 AA.

XX

AC AAW22065;
 XX
 DT 07-OCT-1997 (first entry)
 XX
 DE Chick fringe A (radical) protein.
 XX
 KM Fringe A: radial fringe; fate specification; neural tube;
 KM apical ectodermal ridge; cell proliferation; vasculature;
 KM atherosclerosis; tumour; wound healing; therapy.
 XX
 OS Gallus sp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 27..28
 FT /note= "a frameshift between codons 27 and 28
 FT of the fringe A gene results in fringe
 FT A protein consisting of aa1-27 of one
 FT open reading frame and aa28 onwards of
 FT another reading frame"
 FT
 FT Misc-difference 160
 FT /note= "conserved Cys residue"
 FT
 FT Misc-difference 171
 FT /note= "conserved Cys residue"
 FT
 FT Misc-difference 189
 FT /note= "conserved Cys residue"
 FT
 FT Misc-difference 253
 FT /note= "conserved Cys residue"
 FT
 FT Misc-difference 357
 FT /note= "conserved Cys residue"
 FT
 FT Misc-difference 366
 FT /note= "conserved Cys residue"
 FT
 PN WO9726276-A2.
 XX
 PD 24-JUL-1997.
 XX
 PF 15-JAN-1997; 97WO-US00849.
 XX
 PR 16-JAN-1996; 96US-0586165.
 XX
 PA (HARD) HARVARD COLLEGE.
 XX
 PI Laufer EM, Orozco OE, Tabin CJ;
 XX
 DR WPI, 1997-385295/35.
 DR N-PSDB; AAT75034 and AAT75036.
 XX
 PT Chicken fringe A and fringe B protein(s) and DNA - involved in
 PT differentiation and development of embryonic tissue, used in tumour
 PT treatment
 XX
 PS Claim 9; Fig 1A-B; 64pp; English.
 XX
 CC Chicken fringe A (or radical fringe) protein (AAW22065) has the
 CC ability to affect neural specification of cell type identity,
 CC proliferative regulation of migratory paths, proliferation of
 CC neural crest (branchial arches), definition of morphological
 CC boundaries (including limb), cell shape decisions, regulation of
 CC apical ectodermal ridge (AER) formation and indirect regulation of
 CC signalling to or activating members of the Delta-Serrate-lag-2
 CC family of transmembrane proteins, which in turn bind to the Notch
 CC receptor and affect the differentiation and development of
 CC embryonic tissue. Their amino acid sequences were deduced from
 CC clones cFR42 (AAT75034) and cFR40 (AAT75035) obtd. from a chick embryo
 CC library. Fringe proteins can be used in the treatment of tumour
 CC pathogenesis, disorders of the vasculature system and in wound
 CC healing, and in assays to identify agents which alter the AER and
 CC production of angiogenic precursors, the formation of the AER and
 CC the subdivisions of the neural tube.
 XX
 SO Sequence 372 AA;

Query Match 7.4%; Score 93.5; DB 18; Length 372;
 Best Local Similarity 25.4%; Pred No. 0.53;
 Matches 52; Conservative 17; Mismatches 81; Indels 55; Gaps 11;
 QY 4 CPSIGVTVLVVVLGIRATPAGGQHYLIRPAPSNDLPLVLIHPDPFPD--PREKDL 61
 DB 12 CPLISTVTAALVLLLLPRGPPAPARR---RPPAPAP-----SRPSKRARAPASDV 61
 QY 62 NEYLNSLGLGHDPPGPMATSPEDPRPGGGGAAGAEPLALV---DQLLRPSGAMPS 118
 DB 62 -----PGDR-GGGSGAAGGGRGVAGSPWPSRRVRMGPPGGSAX 98
 QY 119 EIKGLERFEGIAQKKQRLSKLRKLQMWLW-----SQTFCGVLYANNDLSRFMPRYV 173
 DB 99 E--SLEKQDIFIAVKTRTKHKTRLELLPQTMISRAQGTF--IFTWEDRELRL-----L 149
 QY 174 KVSGCFSSKSCS---VPEGNVCKPS 195
 DB 150 KAGDHMINTCNSAVHTRQALCCGMS 174
 RESULT 13
 AAW18658
 ID AAW18658 standard; Protein; 319 AA.
 XX
 AC AAW18658;
 XX
 DT 23-JUL-1997 (first entry)
 XX
 DE Fragmented human NF-L gene +2 frameshift mutant product.
 XX
 KM Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
 KM detection; antibody; probe; cancer; neoplasia; neurodegenerative;
 KM Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
 KM Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
 KM PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
 KM cardiovascular; rheumatoid arthritis; NF-L; neurofilament-low.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1..319
 FT /note= "X corresponds to a stop codon in the
 FT accompanying DNA file, AAT69794"
 FT
 PN WO9712992-A2.
 XX
 PD 10-APR-1997.
 XX
 PF 02-OCT-1996; 96WO-1B01106.
 XX
 PR 11-JAN-1996; 96US-0009832.
 PR 02-OCT-1995; 95GB-0020080.
 XX
 PA (ROYA-) ROYAL NETHERLANDS ACADEMY OF SCI.
 PA (UYUO-) UNIV ROTTERDAM ERASMUS.
 XX
 PI Burbach JPH, Grosveld FG, Van Leeuwen FW;
 XX
 DR WPI, 1997-226235/20.
 DR N-PSDB; AAT69793.
 XX
 PT Use of mutant genes having frameshift mutation(s) - for developing
 PT probe, for the diagnosis, prevention and treatment of associated
 PT diseases, e.g. cancer or neurodegenerative disease
 XX
 PS Claim 22; Fig 7; 123pp; English.
 XX
 CC AAW18657 and AAW18658 are +1 and +2 frameshift mutations, respectively,
 CC of a sequence comprising fragments of the coding sequence of the
 CC human neurofilament subunit NF-Low (NF-L) gene corresponding to
 CC nucleotides (nt) 1-420 followed by nt 901-1440 of the wild-type NF-L

CC gene. These regions of the gene contain GAGAG motifs. Frameshift mutants
CC of the tau, ubiquitin, apolipoprotein E, microtubule-associated protein
CC 2 (MAP-2), neurofilament subunit L, M and H and amyloid A4 genes are
CC claimed. All these genes share a common GAGAN motif (N= A, G, C or T),
CC which is the site of common GA dinucleotide deletion(s) that cause
CC neurodegenerative disorders. Antigenic peptides used for the production
CC of antibodies, and small nucleic acid sequences derived from frameshift
CC mutants are used in the diagnosis, prevention and treatment of cancer
CC and neurodegenerative diseases, e.g. Parkinson's disease, Alzheimer's
CC disease, Down's syndrome, frontal lobe dementia (Pick's disease),
CC progressive supranuclear palsy (PSP), amyotrophic lateral sclerosis,
CC Huntington's disease, multiple sclerosis, and other degenerative
CC diseases such as cardiovascular disease and rheumatoid arthritis.

XX Sequence 319 AA;

Query Match 7.3%; Score 92.5; DB 18; Length 319;

Best Local Similarity 23.8%; Pred. No. 0.55;

Matches 66; Conservative 22; Mismatches 110; Indels 79; Gaps 16;

QY 5 PSIGVTLVALVVVGLRATPAGQGHVLRHAPSDNLP-LVDLIEHPDPIFDPKEDLNE 63

Db 27 PLAGVPXSTAGCVTRRINRG-----RTAERTQPSILPLSPCLP-SP----- 72

QY 64 TLLRSLGHHYDGFMA-----TSPED-----RPGGGGAAGABDLAELDQ 106

Db 73 -----GSHRRRGAAPAAQXVPSATSRTRPTSGATWRRPGCISACAAATAPHAQLTQ 125

QY 107 LLRQR-----PSGAMSEIKGLEFSEGLAQKKQRLSKLRKLQWLW---SQTCPVL 158

Db 126 ATERRCLRPCPCAAAXWTKSL-FXRKCTKRSPNCRRSSTRSPRWTPXPSPT-PPR 183

QY 159 YAWNDLGSFWP-----RYVKVGCFSKR-----SCSVP-----EGMVC---KPS 195

Db 184 SRTSARSTSWPPTCTRLANGRAASRCXPRAPPTPTPCAPTRCRRAVVCSPRW 243

QY 196 KSVH--LTVLRWRCRRGRCGWIPQIPIISECK 230

Db 244 KSKHAGAXMKWRSSCR-----SWRTSRTPTSALCRC 275

RESULT 14

AAAB10043

ID AAB10043 standard; Protein; 538 AA.

XX AAB10043;

XX 02-NOV-2000 (first entry)

XX MMLV gag protein.

XX Glycoprotein; gag gene; pol gene; GP-1; GP-2; anti-HIV; cytostatic;

XX gene therapy; treatment; infectious disease; HIV; AIDS; neoplasm;

XX carcinoma; melanoma.

XX Moloney murine leukemia virus.

XX EP1006196-A2.

XX 07-JUN-2000.

XX 25-NOV-1999; 99EP-0250415.

XX 26-NOV-1998; 98DE-1056463.

XX (PETT-) PETTE INST HEINRICH.

XX Von Laer MD;

XX WPI; 2000-378268/33.

XX New retroviral packing cell useful as pharmaceutical carrier in gene

PT therapy for treatment of HIV and neoplasms, comprises retroviral genes

PT and glycoproteins -

XX Disclosure; Page 31-32; 69pp; German.

XX This invention describes a novel retroviral packing cell (I), comprising
CC the retroviral genes gag, pol and glycoproteins GP-1 and GP-2 of the LCMV
CC coding gene gp, or a part of these. The products of the invention have
CC anti-HIV and cytostatic activity and can be used for gene therapy. (I) is
CC useful for in vitro infection of cells, especially hematopoietic stem
CC cells, for expression of transgenes in cells and as a pharmaceutical
CC carrier for gene therapy. (II) is therefore useful in the treatment of
CC infectious diseases (HIV or AIDS), neoplasms (carcinomas), melanomas and
CC other diseases. This sequence represents the Moloney murine leukemia
CC virus (MMLV) gag protein which is described in the method of the
XX invention.

XX Sequence 538 AA;

Query Match 7.3%; Score 92; DB 21; Length 538;

Best Local Similarity 26.5%; Pred. No. 1.3;

Matches 45; Conservative 12; Mismatches 45; Indels 68; Gaps 9;

QY 21 RATPAGGQHY-----LHIRPAP-----SDN-LPLVDLI-EHPDPIFDPKEDLNETLLRSL 69

Db 122 RSTPPRSSLYPALTPSLGAKPKQVLSGGLIDLLTDEDPYRDP----- 169

QY 70 LGGHYDPGFMATPPEDRPGGGGA--AGGAEDLAELDQLLR-----PSGAMPS 118

Db 170 -----PPSDRDGNGGEATPAGEADPPPMASRLRGRRPPVADSTTSQAFPL 217

QY 119 EIKGLEFSEGLAQKKQRLSKLRKLQWLWSQTFPCVLYAWNDLGSRF 168

Db 218 RAGG-----NGQLQWPFSS--DLYNKNNNPSPF 245

RESULT 15

AAAY17946

ID AAAY17946 standard; Protein; 648 AA.

XX AAAY17946;

XX 03-AUG-1999 (first entry)

XX MoMLV gag gene product.

XX Viral packaging signal; chimeric; type C retrovirus; gag gene; MoMLV;

XX murine VL30; retroviral vector; biotechnology; pharmaceutical;

XX gene therapy; Moloney murine leukemia virus.

XX Moloney murine leukemia virus.

XX WO9925862-A2.

XX 27-MAY-1999.

XX 19-NOV-1998; 98WO-US24667.

XX 19-NOV-1997; 97US-0066148.

XX (NATU-) NATURE TECHNOLOGY INC.

XX Guoping X, Hodgson CP, Zink MA;

XX WPI; 1999-347491/29.

XX N-PSDB; AAX77178.

XX New chimeric packaging signal useful in gene therapy

XX Disclosure; Fig 4; 61pp; English.

XX The invention relates to a new chimeric viral packaging signal that
CC comprises: (i) essential packaging nucleic acid sequence, from a
CC mammalian type C retrovirus, functionally joined to (ii). (ii) at least

